MATHEMATICS OF STATISTICS

PART TWO

BY
JOHN F. KENNEY
University of Wisconsin

SEVENTH PRINTING

NEW YORK
D. VAN NOSTRAND COMPANY, INC.
250 FOURTH AVENUE

First Published May 1939
Reprinted July 1941, September 1942
March 1944, February 1946, March 1947
June 1947

Printed in United States of America

Produced by
TECHNICAL COMPOSITION CO.
BOSTON

Matri Meae

in Signum

Gratitudinis

PREFACE

There are two quite distinct aspects or levels of mathematical statistics. The one involves elementary mathematics and the methodologies serve descriptive purposes. These fundamentals are set forth in Part I. The other aspect is essentially mathematical in character and the methodologies are developed for inferential purposes. It cannot be made elementary by its very nature because the problems are so difficult that powerful mathematical tools are necessary to provide solutions of the problems.

In recent years great advances have been made in statistical theory. Methods of formulating and testing hypotheses have been systematically developed and a sound basis for statistical inference has replaced older methods involving the intuitive notions of "probable error." In this book I have elected to include some of the classical theory and some of the simpler concepts and techniques of the modern theory. In short, I have made a sustained effort to write an up-to-date text which will serve to prepare the student for the really mathematical part of the theory of statistics. A knowledge of elementary probability theory, calculus, and determinants is presupposed. It is also understood that the student is familiar with the rudiments of statistics such as are given in Part I. However, if no preliminary course in statistics has been studied, mature students should be able to acquire the essential definitions and concepts in a rapid survey of Part I.

Of the books which have been particularly useful in preparing the manuscript, I would name the following: Camp's *Mathematical Statistics*, Fisher's *Statistical Methods For Research Workers*, Fry's *Probability And Its Engineering Uses*, Rietz's *Mathematical Statistics*, and Wilks' *Statistical Inference*. I have also derived much help from certain papers in the literature by Professors Carver, Jackson, Rider, and Rietz. Specific reference to these papers is made in the text. A reference list of pertinent books and papers is given at the end of each of the last three chapters. It is recommended that some of these be available to the student for supplementary study in connection with this text.

I am indebted in a very special way to Professor Allen T. Craig: for material taken in his lectures, for a keenly critical reading of the manuscript, and for assistance with the proof reading. Without his help and encouragement Part II could hardly have been written.

Finally, I wish to express my appreciation to D. Van Nostrand Company, Inc., for their efficient and cheerful coöperation in the manufacture of both parts of the book.

John F. Kenney

Evanston, Illinois.
April, 1939

CONTENTS

CHAPTER I

PROBABILITY AND ITS RELATION TO STATISTICAL THEORY. THE BERNOULLI DISTRIBUTION. APPROXIMATIONS BY MEANS OF THE NORMAL CURVE AND POISSON EXPONENTIAL

CHAPTER II

SOME USEFUL INTEGRALS AND FUNCTIONS

CHAPTER III

GENERAL CONCEPTS OF DISTRIBUTION FUNCTION OF A CONTINUOUS VARIABLE. GENERALIZED FREQUENCY CURVE

CHAPTER VII
SMALL OR EXACT SAMPLING THEORY

CHAPTER VIII
A. THE χ^2 DISTRIBUTION
B. STATISTICAL INFERENCE

APPENDIX

MATHEMATICS OF STATISTICS

CHAPTER I

PROBABILITY AND ITS RELATION TO STATISTICAL THEORY. THE BERNOULLI DISTRIBUTION. APPROXIMATIONS BY MEANS OF THE NORMAL CURVE AND POISSON EXPONENTIAL FUNCTION

1. Importance. The subject of probability deals with one of the most interesting branches of modern mathematics and is becoming conspicuous for its applications in many fields of learning. This subject is of fundamental importance, not only in the theory of insurance and statistics, but also in various branches of the biological and physical sciences. The following quotations from contemporary writers indicate the importance of probability theory in the philosophy of modern science.

It was, I think, Huxley who said that six monkeys, set to strum unintelligently on typewriters for millions of millions of years, would be bound in time to write all the books in the British Museum. If we examined the last page which a particular monkey had typed, and found that it had chanced, in its blind strumming, to type a Shakespeare sonnet, we should rightly regard the occurrence as a remarkable accident, but if we looked through all the millions of pages the monkeys had turned off in untold millions of years, we might be sure of finding a Shakespeare sonnet somewhere amongst them, the product of the blind play of chance. . . .

These and other considerations have led many physicists to suppose that there is no determinism in events in which atoms and electrons are involved singly, and that the apparent determinism in large-scale events is only of a statistical nature. When we are dealing with atoms and electrons in crowds, the mathematical law of averages imposes the determinism which physical laws fail to provide. . . . We can only speak in terms of probabilities.

— *The Mysterious Universe*, Sir James Jeans.

In order to understand the nature of knowledge about social and economic life, it is necessary to know something about the theory of probability; because knowledge in these fields, in general, is essentially indeterminate knowledge. There are two fundamental ideas which need to be grasped in order to understand the social sciences. The first idea is that all science is philosophical. . . . The

time honored aim of philosophy has been to discover and interpret (to the extent possible to the human mind) the characteristics of nature. By nature is meant all things, material and psychic, external to man, and man himself. In many fields the minds of men have penetrated into the mysteries of nature and have produced knowledge concerning them. In the physical aspects (both external to man and in man) great progress has been made towards the attainment of apparently precise knowledge, within certain definite limits; while in the field of the psychic the progress has been towards increasing the probabilities of truth of a great variety of hypotheses. But it is characteristic of the psychic aspects of knowledge that the facts in those fields are indeterminate, not precise, and apparently dynamic. Even in the physical and chemical world, the discoveries of recent years have emphasized a great realm of indeterminacy, particularly when confronting great velocities and infinitely small particles within the atom. Thus the second idea to grasp is that in all fields of knowledge, even the physical, beyond the limited range of relatively precise knowledge accumulated by man, there is a vast frontier of speculation. It has been the function of scientific method — the new tool of philosophy — to penetrate ever deeper into this realm of speculative knowledge. Primarily this has been made possible by the development of the theory of probabilities.

— *Elementary Statistics*, James G. Smith.

There exist in nature systems of chance causes which operate in such a way that the effects of these causes can be predicted — by making use of customary probability theory in which objective probabilities in the limiting statistical sense are substituted for the mathematical probabilities.

— *Economic Control of Quality of Manufactured Product*, W. A. Shewhart.

It appears likely that the further development of the theory of probability in the next few decades may turn out to be a major chapter in the history of science.

— *Science*, January 18, 1929.

The great extension in the use of statistics in the last two decades has been associated with and largely made possible by mathematical developments based upon the theory of probability.

— Harold Hotelling, *Journal American Statistical Association*, March Supplement, 1931

2. Definitions. Inasmuch as the subject of probability plays an important role in certain phases of statistical theory, we will now consider some of the fundamental principles of this subject. It will be convenient to divide the subject into two classes, and speak of *a priori* and *empirical* probability.

(*a*) *A priori probability*. If all the ways of obtaining successes and failures can be analyzed into s possible mutually exclusive ways each of which is equally likely, and if x of these ways give successes, the probability of success in a single trial is x/s.

A priori probability is concerned with that class of problems in which a full knowledge of the conditions affecting the event in ques-

tion is known beforehand. In other words, the problem may be set up and solved abstractly. Thus the following problems are questions of *a priori* probability: A box contains 4 white and 6 red billiard balls. What is the probability that two drawn will be of the same color? A coin is to be tossed 7 times. What is the probability that heads will turn up at least 3 times? A sample of telephone receivers is to be taken from a case containing 100 telephone receivers of which 20 are known to be defective. What is the probability that the sample will contain exactly 2 defectives?

There is another class of events in which it is impossible or impractical to enumerate all of the equally likely ways in which the event in question may succeed or fail. When this is the case it is necessary to estimate the probability by trial and observation. Thus we have

(*b*) *Empirical probability.* If it is observed that an event has occurred x times among s trials the ratio x/s is called the relative frequency of success. The limit* of the ratio x/s as s is taken indefinitely large is called the probability of success in a single trial. In symbols we have

$$\lim_{s \to \infty} \frac{x}{s} = p.$$

In statistical applications the limit of x/s cannot in general be determined, but an observed relative frequency (s large) often provides a valuable estimate of the underlying probability assumed in the definition. For example, according to the American Experience

*The student familiar with the theory of limits will realize that a rigorous proof that a probability p exists as the limit of x/s as s increases would require us to show that, Given an $\epsilon > 0$, then there exists a number N such that

$$\left| \frac{x}{s} - p \right| < \epsilon \qquad \text{for all } s \geq N.$$

It is of course obvious that we cannot prove the existence of this limit because we cannot be sure that the difference $|x/s - p|$ will become and remain, as s increases, less than any assigned positive number, no matter how small. For example, after throwing a coin 10,000 times it is possible to get a run of all heads in the next 1000 throws. In this connection Rietz says: "That the limit exists is an empirical assumption whose validity cannot be proved, but experience with data in many fields has given much support to the . . . usefulness of the assumption." (*Mathematical Statistics*, p. 8.)

We can, however, prove that the probability approaches certainty that x/s will approach p as a limit as s is indefinitely increased. (See § 7.)

Mortality Table, out of 57,917 persons living aged 60, there are 1546 who die during the following year. Therefore, the relative frequency 1546/57,917 = .026693 is taken by insurance companies as the probability that a person aged 60 will not survive another year.

3. Theorems. We will now review from algebra certain elementary formulas and theorems leading to the use of probability theory in statistical problems. We will begin with the subject of permutations and combinations.

A *permutation* is an arrangement of all or part of a set of things. A *combination* is a group of all or part of a set of things. A different permutation may be obtained by changing either the items or their order but a different combination may be obtained only by changing one or more of the items in the group.

Theorem I. *The number of permutations of n different things taken r at a time is denoted by the symbol $P(n, r)$ and given by*

$$P(n, r) = n(n-1)(n-2) \cdots (n-r+1).$$

Corollary. *If the n items are not all different, there being n_1 of type T_1, n_2 of type T_2, $\cdots$, n_k of type T_k, then the number of distinct permutations of the n items taken n at a time is*

$$\frac{n\,!}{n_1\,!\,n_2\,! \cdots n_k\,!}$$

where $\sum_1^k n_i = n$. *The symbol $n\,!$, read "factorial n," is defined by*

$$n\,! = n(n-1)(n-2) \cdots 3 \cdot 2 \cdot 1.$$

Theorem II. *The number of combinations of n different things taken r at a time is denoted by $C(n, r)$ and given by*

$$C(n, r) = \frac{n(n-1) \cdots (n-r+1)}{r\,!} = \frac{n\,!}{r\,!\,(n-r)\,!}.$$

It will be understood that $C(n, r)$ equals zero when $r > n$ and equals one when $r = n$.

Theorem III. *The total number of combinations of n different things taken $1, 2, \cdots$, or n at a time is $2^n - 1$.*

Proof. The formula for $C(n, r)$ is the coefficient of the $(r+1)st$ term in the binomial expansion $(x+y)^n$. Thus,

$$(x+y)^n = x^n + C(n, 1)x^{n-1}y + C(n, 2)x^{n-2}y^2 + \cdots + C(n, r)x^{n-r}y^r + \cdots + y^n.$$

If we let $x = y = 1$, this becomes

$$2^n - 1 = C(n, 1) + C(n, 2) + \cdots + C(n, r) + \cdots + C(n, n).$$

Events of a set are said to be *mutually exclusive* if the occurrence of any one of them on a particular occasion excludes the occurrence of any other on that occasion. They are said to be *independent* or *dependent* according as the occurrence of any one of them does not or does affect the occurrence of others in the set.

If p is the probability that an event will happen in a single trial and q is the probability that the event will fail (to happen) in a single trial, then $p + q = 1$ and unity is the symbol for certainty.

Theorem IV. *The probability that one or other of a set of mutually exclusive events should happen when all of them are in question is the sum of the probabilities for the separate events.*

Theorem V. *The probability that all of a set of independent events will happen on a given occasion when all of them are in question is the product of the probabilities for the separate events.*

Theorem VI. *Suppose the events are dependent. Let p_1 be the probability for the happening of a first event E_1 and p_2 be the probability for the occurrence of a second event E_2 after E_1 has happened. Then the probability that both events will happen in the order named is p_1p_2. The procedure may be extended in an obvious manner to any finite number of events.*

4. Supplementary Reading. It is suggested that the student look up the proofs of the above theorems in any college algebra text and review the discussions presented there.

For the more advanced student the following references are recommended. Some of the early chapters of the books may also be read with profit by the beginning student.

Books:

The Mathematical Theory of Probabilities — Arne Fisher.
Probability — Coolidge.
Mathematical Statistics — Rietz.
Choice and Chance — Whitworth.
Probability and Its Engineering Uses — Fry.
Elements of Probability — Levy and Roth.
Introduction to Mathematical Probability — Uspensky.

Papers:

Fundamental Concepts in the Theory of Probability — Fry, American Mathematical Monthly, vol. 41, 1934, p. 207.
On the Foundations of the Theory of Probability — Struik, Philosophy of Science, vol. 1, no. 1, January, 1934.

Problems

1. Prove both algebraically and verbally that
(a) $P(n, r) = C(n, r)P(r, r)$, (b) $C(n, r) = C(n, n - r)$.
2. From among nine men A, B, C, D, E, F, G, H, I, a committee of four men will be chosen. The nine names will be written on nine separate cards and four cards drawn at random one at a time from a box.
   (a) In how many different ways may the four cards come out? *Ans.* 3024.
   (b) How many different committees are possible not including the man A? *Ans.* 70.
3. Consider the word "introduce."
   (a) In how many of the possible arrangements of all its letters will there be a consonant in the first place? *Ans.* 201,600.
   (b) From its letters how many four letter permutations consisting of three vowels and one consonant can be formed? *Ans.* 480.
   (c) If five of its letters are selected at random what is the probability that two are vowels and three are consonants? *Ans.* 10/21.
4. On a table there are four different biographies with brown backs and seven different novels with red backs.
   (a) If all of the books are placed upright in a row on a shelf, in how many different ways may they be arranged so that the orders of the colors are different? *Ans.* 330.
   (b) In how many different ways may two of the biographies and three of the novels be selected and arranged on the shelf so that the orders of the books are different? *Ans.* 25,200.
5. In a box there are five red billiard balls with the numbers 1, 2, 3, 4, 5, painted on them (one on each ball), and three white billiard balls with the numbers 1, 2, 3, similarly painted on them. From the box a man draws two balls at random.
   (a) What is the probability that one of the balls drawn is white and the other is red? *Ans.* 15/28.
   (b) What is the probability that the two balls drawn have either the same color or the same number? *Ans.* 4/7.
6. A bag contains four white, five red, and six black balls. Three are drawn at random. Find the probability that (a) no ball drawn is black, (b) exactly two are black, (c) all are of the same color.
7. An urn contains four white and five black balls. Three balls are drawn at random and replaced by green balls. If then two balls are drawn at random, what is the probability that they are both of the same color? *Ans.* 29/108.
8. Write out the expressions for $C(n - 1, 2)$; $C(n - 1, 3)$; $C(s, x)$.
9. (a) Show that

$$C(s-1, x-1) = \frac{(s-1)(s-2)\cdots(s-x+1)}{(x-1)!} = \frac{(s-1)!}{(x-1)!(s-x)!}.$$

   (b) What is the value of the above expression when $x = 1$?

10. Write in expanded form:

(a) $\sum_{x=0}^{s} C(s, x)$.

(b) $\sum_{x=1}^{s} C(s-1, x-1)$.

11. Twelve cards have been dealt, six down, and the other six showing a jack, two kings, a seven, a five, and a four. What is the probability that the next card will be a four or less? (*National Mathematics Magazine*, vol. XIII, no. 2, p. 94.)

12. From an urn containing ten balls, numbered from one to ten, balls are drawn, one by one and placed in a row of holes, numbered from one to ten, each ball being placed in the proper hole. What is the probability that there will not be an empty hole between two filled ones at any time of the drawing? (*American Mathematical Monthly*, vol. 45, no. 9, p. 635.) *Ans.* 2/14,175.

5. Repeated Trials. We now consider a theorem which is very important both in the theory of probability and its applications in statistics.

Theorem VII. *Let p be the probability that an event will happen in a single trial, and $q = 1 - p$ the probability that the event will fail in a single trial. Then the probability P that the event will happen exactly x times in s trials, during which p remains constant, is given by the $(x + 1)$st term of the binomial expansion:*

$$(1) \qquad (q + p)^s = q^s + C(s, 1)pq^{s-1} + C(s, 2)p^2q^{s-2} + \cdots + C(s, x)p^xq^{s-x} + \cdots + p^s.$$

Proof. By Theorem V, the probability that the event will happen x times and fail the other $s - x$ times in any specified order is p^xq^{s-x}. But the number of ways in which the order may be specified is $C(s, x)$ or $C(s, s - x)$. These ways are equally likely and mutually exclusive. Therefore, by Theorem IV the required probability is $C(s, x)p^xq^{s-x}$. We recognize this expression as the $(x + 1)$st term of the binomial expansion of $(q + p)^s$.

COROLLARY 1. *The probability that the event will happen at most x times in s trials is the sum of all those terms of* (1) *in which the exponent of p is equal to or less than x.*

COROLLARY 2. *The probability that the event will happen at least x times in s trials is the sum of all those terms of* (1) *in which the exponent of p is equal to or greater than x.*

Proofs. By Theorem IV, the probability that the event will happen at most x times is the sum of the probabilities that it will happen

$0, 1, 2, 3, \cdots, x$ times. Similarly, the probability that the event will happen at least x times is the sum of the probabilities that it will happen $s, s-1, s-2, \cdots, x$ times.

Problems

1. A dust storm contains particles of two kinds identical except as to color, brown and yellow particles existing in the ratio 3 : 2. If five particles of this dust enter my eye at random determine the probability that two of them are brown and the other three are yellow. (See *American Mathematical Monthly*, vol. 41, no. 5, May 1934.)
2. Six coins are tossed once, or what amounts to the same thing, one coin is tossed six times. Find the probability of obtaining heads
   (*a*) exactly three times
   (*b*) at most three times
   (*c*) at least three times
   (*d*) at least once.
3. (*a*) What is the probability of throwing seven in a single toss of two dice?
   (*b*) In six tosses of two dice find the probability of throwing seven at least once.
4. Toss six coins 64 times and record the number of times heads appear 0, 1, 2, 3, 4, 5, 6 times. (Instead of tosses, the coins may be shaken in a box.) Compare the resulting distribution of frequencies with the terms of the expansion of $64(\frac{1}{2} + \frac{1}{2})^6$.
5. A bag contains white and black balls in the proportion 2 : 3. Let the probability of drawing a white ball be called a success. Three balls are drawn separately and after each drawing the ball is returned to the bag and thoroughly mixed with the others so that the fundamental probability of success remains constant during the trials. Find the probabilities of 0, 1, 2, 3 successes. If this experiment were repeated 125 times what is the theoretical frequency of each of the possible number of successes?
6. Show that equation (1) may be written:

$$(q + p)^s = \sum_{x=0}^{s} \frac{s\,!}{x\,!\,(s - x)\,!} p^x q^{s-x}.$$

7. Show that

$$\sum_{x=1}^{s} \frac{(s - 1)\,!}{(x - 1)\,!\,(s - x)\,!} p^{x-1} q^{s-x} = (q + p)^{s-1} = 1.$$

8. (*a*) Find the values of $C(18, x)$ for $x = 0$ to $x = 18$ inclusive. (To the instructor: Pascal's Triangle provides a simple scheme for constructing a table of binomial coefficients.)
   (*b*) Evaluate $2^x/3^{18}$ for $x = 0$ to $x = 18$ inclusive.
   (*c*) Show that $(\frac{1}{3} + \frac{2}{3})^{18}$ may be written

$$\sum_{x=0}^{18} f(x) \text{ where } f(x) = C(18, x)2^x/3^{18}.$$

   (*d*) Using the results of (*a*) and (*b*), find the values of $f(x)$ for $x = 0$ to $x = 18$. Save your results for future reference.

6. Relative Frequencies from Dichotomous Samples. Suppose a sample of s individuals from the same population is divided into two groups according as they have a certain attribute or not. Such a division is said to be *dichotomous.* Out of s individuals we find that x individuals have the attribute in question and $s - x$ do not, it being possible for x to take any integral value from 0 to s inclusive. The attribute in question is frequently called the " event " and its occurrence is called a " success." The ratio x/s is called the relative frequency of success.

Many illustrations of relative frequency come readily to mind. Out of 100 throws of a coin we may have noted 45 heads. From a group of school children, taken at random, we may find 55 boys. Or again, we might make a certain disease of children the basis of a dichotomy. Out of 100 fifth grade school children we may find that 27/100 is the relative frequency of the occurrence of measles.

7. Theorem of Bernoulli. The theorem of Bernoulli describes the approach of the relative frequency x/s to the underlying constant probability p as s increases. The theorem may be stated as follows:

Theorem VIII. *In a set of s trials in which the chance of success in each trial is a constant p, the probability P approaches unity that the relative frequency x/s will approach p as a limit as s increases indefinitely.**

Observe that this is a weaker statement than saying that p is the limit of x/s as the number of trials increases indefinitely. Another way of stating the theorem is as follows: The probability $Q = 1 - P$ of the difference $(x/s - p)$ being numerically as large as any assigned positive number ϵ will approach zero as a limit as s increases indefinitely.

The theorem is the basis for our definition of empirical probability. It is often regarded as a fundamental theorem of mathematical statistics because of the common use of x/s (s large) as a close approximation to the probability p.

8. Binomial Description of Frequency. The terms of $(q + p)^s$ are the theoretical relative frequencies for a dichotomous situation. If we take N sets of s trials the theoretical absolute frequencies are given by the terms of $N(q + p)^s$ when N is chosen so that these terms are integers. It follows that N is merely a proportionality factor.

* A proof is given in Chapter VI, §10.

Hence we may say that if in a single trial the probability of an event occurring is p and the probability of its not occurring is q, then if a sample of s trials is taken, the frequencies with which the event occurs $0, 1, 2, 3, \cdots, s$ times are proportional to the terms of the point binomial $(q + p)^s$. This was the first theoretical distribution to be established and a discussion of it is given in *Ars Conjectandi* by J. Bernoulli which was published posthumously in 1713. A distribution of discrete variates with frequencies proportional to the terms of (1) is frequently referred to as a Bernoulli distribution.

In the *Carus Monograph on Mathematical Statistics* (p. 23) Professor Rietz explains the applications and limitations of (1) in practical statistics as follows:

> Such a distribution . . . serves as a norm for the distributions of relative frequencies obtained from some of the simplest sampling operations in applied statistics. For example, the geneticist may regard the Bernoulli distribution (1) as the theoretical distribution of the relative frequencies x/s of green peas which he would obtain among random samples each consisting of a yield of s peas. The biologist may regard (1) as the theoretical distribution of the relative frequencies of male births in samples of s births. The actuary may regard (1) as the theoretical distribution of yearly death rates in samples of s men of equal ages, say of age 30, drawn from a carefully described class of men. In this case we specify that the samples shall be taken from a carefully described class of men because the underlying assumptions involved in (1) do not permit a careless selection of data. Thus, it would not be in accord with the assumptions to take some of the samples from a group of teachers with a relatively low rate of mortality and others from a group of anthracite coal miners with a relatively high rate of mortality. . . .
>
> The expression "simple sampling" is sometimes applied to drawing a random sample when the conditions for repetition just described are fulfilled. In other words, simple sampling implies that we may assume the underlying probability p of formula (1) remains constant from sample to sample, and that the drawings are mutually independent in the sense that the results of drawings do not depend in any significant manner on what has happened in previous drawings.

9. Graphical Representation. A binomial distribution may be represented graphically by a histogram. This is accomplished by constructing rectangles centered at $x = 0, 1, 2, \cdots, s$ with heights proportional to the terms of the binomial. The different "successes" denoted by x are the variates, and the corresponding terms of the binomial are the theoretical relative frequencies.

Since the values of x constitute a discrete series it might seem more logical to represent the relative frequencies by ordinates instead of rectangles. However, since the base of each rectangle is unity the

number representing its height is also its area, and the representation by areas will be useful in our work. In a case like this the frequencies are said to be "loaded" on the ordinates at the mid-points of the class intervals.

If we are thinking of relative frequencies or probabilities the sum of all the rectangles is unity, whereas if we are thinking of absolute frequencies the total area of the histogram is N. Thus if six coins are tossed 64 times the theoretical absolute frequencies are given by the terms of $64(\frac{1}{2} + \frac{1}{2})^6$. These are 1, 6, 15, 20, 15, 6, 1 and their sum is 64.

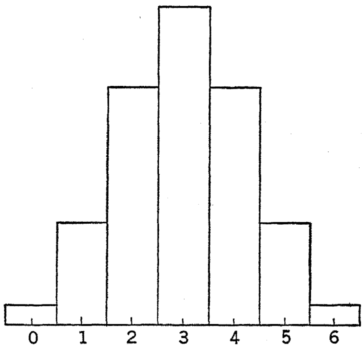


FIG. 1. HISTOGRAM of $(\frac{1}{2} + \frac{1}{2})^6$

10. The Mean and Standard Deviation. We have shown that the terms of $N(q + p)^s$ give the expected frequency of success (with respect to an attribute or character) in drawing N samples of s items in each sample, where p is the probability of a success. We now propose to characterize the distribution of expected frequencies by finding the usual moments. In this procedure we may consider the relative frequencies given by the terms of $(q + p)^s$ because the absolute frequencies are proportional to these terms, N being the proportionality factor. It will be convenient to evaluate first the ν's, taking the position of the first term as origin.

By definition

$$\bar{x} = \nu_1 = \frac{\sum xf(x)}{\sum f(x)}$$

where x refers to the number of successes and $f(x)$ refers to the corresponding probabilities which are of course the theoretical relative frequencies. Table 1 shows the appropriate frequency table. It is obvious that the sum of the second column is unity. To sum the third column we factor out sp, obtaining

$$sp[q^{s-1} + (s-1)pq^{s-2} + C(s-1,\, 2)p^2q^{s-3} \\ + \cdots + C(s-1,\, x-1)p^{x-1}q^{s-x} + \cdots + p^{s-1}]$$

which may be written $sp(q + p)^{s-1} = sp$. Hence, we have that the *mean number of successes in s trials is* $\bar{x} = sp$, *where p is the probability*

TABLE 1

x	f	xf
0	q^s	0
1	spq^{s-1}	spq^{s-1}
2	$\frac{s(s-1)}{2!}p^2q^{s-2}$	$s(s-1)p^2q^{s-2}$
3	$\frac{s(s-1)(s-2)}{3!}p^3q^{s-3}$	$\frac{s(s-1)(s-2)}{2!}p^3q^{s-3}$
—	. . . .	. . . .
x	$\frac{s(s-1)\cdots(s-x+1)}{x!}p^xq^{s-x}$	$\frac{s(s-1)\cdots(s-x+1)}{(x-1)!}p^xq^{s-x}$
—	. . . .	. . . .
s	p^s	sp^s
Totals	$\sum f(x) = (q+p)^s = 1$	$\sum xf(x) = sp(q+p)^{s-1} = sp$

of success in any trial. This result is often called the " mathematical expectation " or the " expected value " of x.

Table 1 assists our intuitions but logically it is unnecessary. We could have proceeded as follows:

$$\nu_1 = \sum_{x=0}^{s} \frac{s!}{x!\,(s-x)!}p^xq^{s-x}x \div \sum_{x=0}^{s} \frac{s!}{x!\,(s-x)!}p^xq^{s-x}.$$

We observe that the divisor is unity and in the dividend we can divide x into x !. So,

$$\nu_1 = \sum_{x=1}^{s} \frac{s!}{(x-1)!\,(s-x)!}p^xq^{s-x}.$$

Factoring out sp, we have

$$\begin{aligned}\nu_1 &= sp\sum_{1}^{s} \frac{(s-1)!}{(x-1)!\,(s-x)!}p^{x-1}q^{s-x}\\ &= sp(q+p)^{s-1}\end{aligned}$$

whence we obtain

(2) $$\bar{x} = sp.$$

We will use this procedure in finding the higher moments. Since

$$\sum f(x) = \sum_0^s C(s, x)p^x q^{s-x} = 1$$

we may omit it from the denominators of the rest of the ν's. By definition then

$$\nu_2 = \sum_0^s \frac{s\,!}{x\,!\,(s-x)\,!} p^x q^{s-x} x^2.$$

Writing $x^2 = x\,(x-1) + x$, we have

$$\nu_2 = \sum_0^s \frac{s\,!}{x\,!\,(s-x)\,!} p^x q^{s-x} x(x-1) + \sum_0^s \frac{s\,!}{x\,!\,(s-x)\,!} p^x q^{s-x} x.$$

This simplifies into

$$s(s-1)p^2(q+p)^{s-2} + sp$$

so that we obtain

$$\nu_2 = s(s-1)p^2 + sp.$$

In order to get σ we must know the second moment about sp. From the relation $\mu_2 = \nu_2 - (\nu_1)^2$ we easily find that

$$\mu_2 = spq$$

whence

$$\sigma = (spq)^{1/2}. \tag{3}$$

Example 1. Find the mean and standard deviation of the binomial $(\frac{2}{5} + \frac{3}{5})^5$ by means of formulas (2) and (3). Verify your results by the usual procedure for computing moments of a frequency distribution.

Solution. Here $p = \frac{3}{5}$, $q = \frac{2}{5}$, $s = 5$. By formulas (2) and (3), $\bar{x} = 3$, $\sigma = 1.095$.

Verification. $(\frac{2}{5} + \frac{3}{5})^5 = \frac{1}{5^5}[32 + 240 + 720 + 1080 + 810 + 243]$.

In finding the moments we may omit the proportionality factor $1/5^5$.

x	f	u
0	32	−3
1	240	−2
2	720	−1
3	1080	0
4	810	1
5	243	2

We find $\sum f = 3125$, $\sum uf = 0$, $\sum u^2 f = 3750$.

Hence
$\bar{u} = 0$, $\bar{x} = x_0 + c\bar{u} = 3$,
$\mu_2 = 1.2$, $\sigma_x = \sigma_u = \sqrt{1.2} = 1.095$.

11. Skewness and Kurtosis. We shall now derive expressions for the third and fourth moments. By definition

$$\nu_3 = \sum_{x=0}^{s} \frac{s\,!}{x\,!\,(s-x)\,!} p^x q^{s-x} x^3.$$

Writing $x^3 = x(x-1)(x-2) + 3x^2 - 2x$, we have

$$\begin{aligned}
\nu_3 &= \sum_{x=0}^{s} \frac{s\,!}{x\,!\,(s-x)\,!} p^x q^{s-x} x\,(x-1)(x-2) \\
&\quad + 3\sum_{x=0}^{s} \frac{s\,!}{x\,!\,(s-x)\,!} p^x q^{s-x} x^2 \\
&\quad - 2\sum_{x=0}^{s} \frac{s\,!}{x\,!\,(s-x)\,!} p^x q^{s-x} x \\
&= s(s-1)(s-2)p^3 \sum_{x=3}^{s} \frac{(s-3\,!)}{(x-3)\,!\,(s-x)\,!} p^{x-3} q^{s-x} \\
&\quad + 3[s(s-1)p^2 + sp] - 2\,sp \\
&= s(s-1)(s-2)p^3 + 3s(s-1)p^2 + sp.
\end{aligned}$$

Similarly, by definition

$$\nu_4 = \sum_{x=0}^{s} \frac{s\,!}{x\,!\,(s-x)\,!} p^x q^{s-x} x^4.$$

Writing $x^4 = x(x-1)(x-2)(x-3) + 6x^3 - 11x^2 + 6x$ and proceeding in a way analogous to that for evaluating ν_3 we obtain

$$\nu_4 = s(s-1)(s-2)(s-3)p^4 + 6\nu_3 - 11\nu_2 + 6\nu_1.$$

Next we desire the moments about the mean, so that we may obtain expressions for skewness and kurtosis. From the relations

$$\begin{aligned}
\mu_3 &= \nu_3 - 3\nu_2\nu_1 + 2\nu_1^3 \\
\mu_4 &= \nu_4 - 4\nu_3\nu_1 + 6\nu_2\nu_1^2 - 3\nu_1^4
\end{aligned}$$

we obtain the quite simple results

$$\begin{aligned}
\mu_3 &= spq(q-p) \\
\mu_4 &= spq[1 + 3(s-2)pq].
\end{aligned}$$

Recalling that $\alpha_r = \mu_r/\sigma^r$ we have finally that

$$\alpha_3 = \frac{(q - p)}{\sqrt{spq}}$$

$$\alpha_4 = \frac{1}{spq} - \frac{6}{s} + 3.$$

We observe that none of these moments are subject to Sheppard's corrections because the assumption that all the values are concentrated at the mid-point of an interval is actually true in the case of a binomial distribution. This is obvious graphically since each frequency is represented by the middle of a rectangle.

12. A Recursion Formula. The moments μ_k of a Bernoulli distribution can be obtained in an elegant manner by means of the recursion formula

$$\mu_{k+1} = pq\left[sk\mu_{k-1} - \frac{d\mu_k}{dq}\right]. \tag{4}$$

We know that $\mu_0 = 1$ and $\mu_1 = 0$, so the formula is to be used for $k \geq 1$. Thus for

$$\begin{aligned}
k = 1, \qquad \mu_2 &= pq(s\mu_0 - 0) \\
&= spq. \\
k = 2, \qquad \mu_3 &= pq[0 - (s - 2sq)] \\
&= spq(2q - 1) \\
&= spq(q - p). \\
k = 3, \qquad \mu_4 &= pq[3s^2pq + s - 6sq + 6sq^2] \\
&= spq[1 + 3spq - 6pq] \\
&= spq[1 + 3(s - 2)pq].
\end{aligned}$$

A simple proof of this formula has been given by A. T. Craig in the *Bulletin of the American Mathematical Society*, vol. 40, pp. 262–264.

To summarize, we have the important characterizing functions of a Bernoulli distribution:

$$\left\{\begin{array}{ll}
\textbf{Mean:} & \bar{x} = sp \\
\textbf{Variance:} & \sigma^2 = spq \\
\textbf{Skewness:} & \alpha_3 = (q - p)/\sigma \\
\textbf{Kurtosis:} & \alpha_4 = 1/\sigma^2 - 6/s + 3 \\
\textbf{Excess:} & \alpha_4 - 3 = (1 - 6pq)/spq.
\end{array}\right.$$

13. Mathematical Expectation. If a variable x may assume any one of a countable set of mutually exclusive values $x_1, x_2, \cdots, x_n$, in such a way that $f(x_i)$, which we take to be single-valued and non-negative, is the probability that x takes the value x_i and such that $\sum_{i=1}^{n} f(x_i) = 1$, then x is called a chance variable and $f(x)$ is defined as the probability function of the discrete variable x. If the mutually exclusive values are 0, 1, 2, 3, $\cdots$, s, an example of such a law of probability is

$$f(x) = C(s, x)p^x q^{s-x}.$$

A frequency distribution whose relative frequencies are given in accord with this law of probability is styled a Bernoulli distribution, as we have already observed.

Let the discrete variable x be subject to the law of probability $f(x)$ and let $g(x)$ be any function of x. The *mathematical expectation* of $g(x)$, denoted by application of the operator E, is then defined to be

$$E[g(x)] = \sum_{i=1}^{n} g(x_i) f(x_i).$$

In particular, if $g(x) = x$ then

$$\begin{aligned} E(x) &= \sum_{i=1}^{n} x_i f(x_i) \\ &= \nu_1 = \bar{x} \end{aligned}$$

is the first moment, per unit frequency, about the origin. More generally, if $g(x) = x^k$, $(k = 1, 2, \cdots)$, then

$$E(x^k) = \sum_{i=1}^{n} x_i^k f(x_i) = \nu_k$$

is the kth moment about the origin. If $f(x) = C(s, x)p^x q^{s-x}$ and $g(x) = x^k$, then

$$E(x^k) = \sum_{k=0}^{s} x^k C(s, x) p^x q^{s-x} \tag{5}$$

defines the moments, about the origin, of a Bernoulli distribution. In particular for $k = 1$, we have

$$E(x) = sp. \tag{6}$$

If $g(x) = (x - sp)^k$ and $f(x) = C(s, x)p^x q^{s-x}$, then

$$E[(x - sp)^k] = \sum_{x=0}^{s} (x - sp)^k C(s, x) p^x q^{s-x}$$

is the kth moment about the mean. For $k = 1$, we see that $E(x - sp) = 0$, and for $k = 2$, we have

$$\begin{aligned}\sigma_x^2 &= E(x - sp)^2 \\ &= E(x^2) - (sp)^2\end{aligned}$$

which we have seen reduces to

$$E(x - sp)^2 = spq. \tag{7}$$

Equations (6) and (7) give the mean and variance with respect to the number of successes x in s trials. In some statistical investigations the data are expressed in terms of percentages or rates. When we may assume a constant probability underlying the frequency ratios obtained from observations we have a binomial distribution as before but on a different scale. Instead of the variable being x it is now x/s. In this case we have

$$E\left(\frac{x}{s}\right) = \frac{1}{s}E(x) = \frac{sp}{s} = p. \tag{8}$$

For the analogous concept relating to the variance we have

$$E\left(\frac{x}{s} - p\right)^2 = \frac{1}{s^2}E(x - sp)^2 = \frac{spq}{s^2} = \frac{pq}{s}. \tag{9}$$

Therefore, we see from (6) and (7) that the *number* of successes per set of s trials is distributed about an expected value of sp with a standard deviation of $(spq)^{1/2}$. From (8) and (9) we see that the *percentage* of successes in a set of s trials is distributed about an expected value of p with a standard deviation of $(pq/s)^{1/2}$.

In probability theory, the standard deviation is often called the standard error. It is important to observe that for a fixed value of p the standard error of x about sp increases as s increases and is proportional to $(s)^{1/2}$, whereas the standard error of x/s about p decreases as s increases, since it is proportional to $(1/s)^{1/2}$.

Exercises

1. Expand the binomial $N(\frac{1}{3} + \frac{2}{3})^s$ for $s = 2$ and $s = 8$. Find the theoretical frequencies in each case by taking N as the smallest number necessary to express the terms of each expansion as integers.
2. Find the mean and standard deviation for each of the above distributions using the appropriate formulas in (4).
3. Find $\bar{x}$, σ, α_3, α_4 for each of the following binomials:

$$(\tfrac{1}{2} + \tfrac{1}{2})^7,\ (\tfrac{1}{6} + \tfrac{5}{6})^4,\ (\tfrac{1}{3} + \tfrac{2}{3})^{18}.$$

4. For a certain binomial distribution
$\sigma = 2.66$, $\alpha_3 = 0.318$. Find p, q, and s.
5. Assume that .04 is the theoretical rate of mortality in a certain age group. Suppose an insurance company is carrying $s = 1000$ such cases. What is the expected dispersion (standard error) in death rates from the theoretical rate $p = .04$? What would it be if $s = 10{,}000$?
6. The value of x for which $C(s, x)p^x q^{s-x}$ is the largest is called the mode of a Bernoulli distribution. Show that the mode is the positive integral value (or values) of x for which

$$sp - q \leq x \leq sp + p.$$

References:
1. *Mathematical Theory of Probabilities* — Fisher, pp. 99–101.
2. *Mathematical Statistics* — Rietz, p. 25.

7. Suppose the law of distribution of the happening of an event in s successive trials is given by the terms of the expansion of

$$(q + p)^s = \sum_{x=0}^{s} C(s, x)p^x q^{s-x} = \sum_{x=0}^{s} P_x.$$

(*a*) If $s = 100$ what values of p and q will make $P_0 = P_1$; $P_9 = P_{10}$?
(*b*) Give approximate values of the P's in (*a*).
8. A bag contains three one dollar bills and four five dollar bills. Three bills are drawn at random. For each one dollar bill withdrawn, three two dollar bills are returned to the bag, and for each five dollar bill that is drawn, a one and a two and a ten dollar bill are returned to the bag. A second drawing of two bills is made. Designate by x and y, respectively, the values of the first and second drawings. (*a*) Give in tabular form the probabilities for each of the possible simultaneous values of x and y. (*b*) Evaluate $E(x)$ and $E(y)$.

Solution. (*a*) The required probabilities are given in the cells of the table on page 19. The marginal totals are denoted by $g(x_i)$ and $h(y_j)$. The fact that $\sum_1^n g(x_i) = 1$ and $\sum_1^m h(y_j) = 1$ is a check on the computations. (*b*) $E(x) = \sum_1^n x_i g(x_i) = 26{,}910/2730 = \$9.86$, $E(y) = \sum_1^m y_j h(y_j) = 18{,}120/2730 = \$6.64$.

9. A bag contains three one dollar bills and two two dollar bills. Two bills are drawn at random. For each one dollar bill drawn two two dollar bills are returned to the bag, while for each two dollar bill drawn a one and a two dollar bill are returned to the bag. A second drawing of two bills is made. Designate by x and y, respectively, the values of the first and second drawings. Give in tabular form the probabilities for each of the possible simultaneous values of x and y. Find $E(x)$ and $E(y)$.
10. For the more advanced student: Read and report on the following article, *Urn Schemata as a Basis for the Development of Correlation Theory* — Rietz, Annals of Mathematics, (2), vol. 21 (1920), p. 306.

$y \backslash x$	3	7	11	15	$h(y_i)$
20	$\frac{1}{35} \cdot 0$	$\frac{12}{35} \cdot 0$	$\frac{18}{35} \cdot \frac{1}{78}$	$\frac{4}{35} \cdot \frac{3}{78}$	$\frac{30}{2730}$
15	$\frac{1}{35} \cdot 0$	$\frac{12}{35} \cdot \frac{3}{78}$	$\frac{18}{35} \cdot \frac{4}{78}$	$\frac{4}{35} \cdot \frac{3}{78}$	$\frac{120}{2730}$
12	$\frac{1}{35} \cdot 0$	$\frac{12}{35} \cdot \frac{7}{78}$	$\frac{18}{35} \cdot \frac{10}{78}$	$\frac{4}{35} \cdot \frac{9}{78}$	$\frac{300}{2730}$
11	$\frac{1}{35} \cdot 0$	$\frac{12}{35} \cdot \frac{2}{78}$	$\frac{18}{35} \cdot \frac{8}{78}$	$\frac{4}{35} \cdot \frac{18}{78}$	$\frac{240}{2730}$
10	$\frac{1}{35} \cdot \frac{6}{78}$	$\frac{12}{35} \cdot \frac{3}{78}$	$\frac{18}{35} \cdot \frac{1}{78}$	$\frac{4}{35} \cdot 0$	$\frac{60}{2730}$
7	$\frac{1}{35} \cdot \frac{36}{78}$	$\frac{12}{35} \cdot \frac{21}{78}$	$\frac{18}{35} \cdot \frac{10}{78}$	$\frac{4}{35} \cdot \frac{3}{78}$	$\frac{480}{2730}$
6	$\frac{1}{35} \cdot 0$	$\frac{12}{35} \cdot \frac{6}{78}$	$\frac{18}{35} \cdot \frac{8}{78}$	$\frac{4}{35} \cdot \frac{6}{78}$	$\frac{240}{2730}$
4	$\frac{1}{35} \cdot \frac{36}{78}$	$\frac{12}{35} \cdot \frac{21}{78}$	$\frac{18}{35} \cdot \frac{10}{78}$	$\frac{4}{35} \cdot \frac{3}{78}$	$\frac{480}{2730}$
3	$\frac{1}{35} \cdot 0$	$\frac{12}{35} \cdot \frac{14}{78}$	$\frac{18}{35} \cdot \frac{20}{78}$	$\frac{4}{35} \cdot \frac{18}{78}$	$\frac{600}{2730}$
2	$\frac{1}{35} \cdot 0$	$\frac{12}{35} \cdot \frac{1}{78}$	$\frac{18}{35} \cdot \frac{6}{78}$	$\frac{4}{35} \cdot \frac{15}{78}$	$\frac{180}{2730}$
$g(x_i)$	$\frac{78}{2730}$	$\frac{936}{2730}$	$\frac{1404}{2730}$	$\frac{312}{2730}$	1

14. Approximating the Binomial with the Normal Curve. If we plot the terms of $(q + p)^s$ as ordinates against the values of $x/\sqrt{s}$ as abscissas and draw the corresponding histogram we find that it approaches a smooth curve as s is taken larger and larger. Thus in Figure 2 (where the vertical sides of the rectangles are omitted since they contribute nothing to the interpretation) we see how the staircase outline of the histogram approaches close to a continuous curve as s is taken larger.

The limiting values of α_3 and α_4 for the binomial as $s \to \infty$ are those of the normal curve. Thus from

$$\alpha_3 = \frac{q - p}{\sqrt{spq}}$$

and

$$\alpha_4 = \frac{1}{spq} - \frac{6}{s} + 3$$

we see that $\alpha_3 \to 0$ and $\alpha_4 \to 3$ as $s \to \infty$. This suggests the possibility of approximating the binomial with the normal curve. As a matter of fact, it can be proved, under certain conditions of approximation, that $(q + p)^s$ approaches the normal curve as a limit as $s \to \infty$. The proof* will not be given here but a word or two about it may be appropriate. In using the normal curve to approximate the binomial we are particularly interested in a range of three or four standard deviations from the mean. This fact suggests the reasonableness of assuming that the number of successes x' above or below sp be considered as the same order of magnitude as σ. This means that $x'/(spq)^{1/2}$ shall remain finite as $s \to \infty$. Now $(spq)^{1/2}$ is of order $(s)^{1/2}$ if neither p nor q is extremely small. Hence the propriety of assuming (in the proof) that $x'/(s)^{1/2}$ shall remain finite. This is the reason for plotting the histograms (Figure 2) in terms of $x/(s)^{1/2}$.

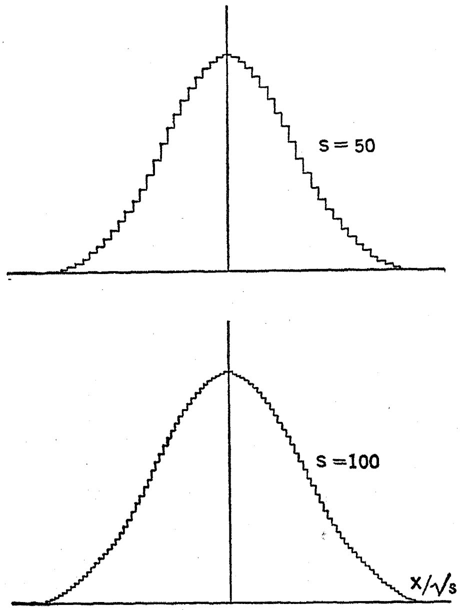


FIG. 2. SHOWING APPROACH OF $(q + p)^s$ TO SMOOTH CURVE AS $s \to \infty$

We may expect, therefore, that the fitted normal curve will give a fair approximation to the binomial except possibly at the extremities of the range. When the terms of the binomial are arranged symmetrically with respect to the mean, that is, when $p = q$, the approximation is rather better than otherwise.

* The following references are recommended:
Mathematical Statistics — Rietz, pp. 32–35.
Probability and Its Engineering Uses — Fry, pp. 207–213.
Annals of Mathematical Statistics, vol. 1, p. 197.

Exercise

Fit a normal curve to the binomial $(\frac{1}{3} + \frac{2}{3})^{18}$. Directions: This binomial may be written

$$\sum_{x=0}^{18} f(x) \qquad \text{where} \qquad f(x) = C(18, x)\frac{2^x}{3^{18}}.$$

(See Problem 8, § 5.) Next recall that the equation of the normal curve is

$$y = \frac{N}{\sigma}\phi(t)$$

where $$\phi(t) = \frac{1}{\sqrt{2\pi}}e^{-t^2/2} \qquad \text{and} \qquad t = \frac{(x - \bar{x})}{\sigma}$$

If we set $N = 1$, $\bar{x} = sp$, and $\sigma = (spq)^{1/2}$ we shall expect that y will give, approximately, the values of $f(x)$ for the various values of x. As in Chapter VI of Part I the following outline is suggested for organizing the computations.

x	t	$\phi(t)$	y	$f(x)$

Construct the histogram and draw the curve. It is suggested that paper ruled "20 to the inch" be used. By comparing the last two columns and also judging from the figure, does the fit seem to be good, even though s is rather small and $q = \frac{1}{2}p$?

The above exercise will help the student appreciate a theorem which will now be introduced. The sum of successive terms of the binomial equals the area of the corresponding rectangles in its histogram. We may obtain an approximation to this sum by finding the area under the fitted normal curve which these rectangles occupy. Graphically, the values $x = 0, 1, 2, \cdots, s$ are the mid-points of the bases of these rectangles. Therefore, if we are summing the terms of the binomial in which x ranges from $x = d_1$ to $x = d_2$, inclusive, the corresponding area under the curve will be from $x = d_1 - \frac{1}{2}$ to $x = d_2 + \frac{1}{2}$. We must convert these values into standard units in order to enter a table of areas of the normal curve. Hence we have the following theorem.

Theorem IX.* *The sum of those terms of the binomial $(q + p)^s$ in which the number of successes x ranges from d_1 to d_2, inclusive, is approximately*

$$Q = \int_{t_1}^{t_2} \phi(t)dt$$

* Sometimes called the De Moivre-Laplace Theorem.

where

$$t_1 = \frac{d_1 - \frac{1}{2} - sp}{\sigma}, \qquad t_2 = \frac{d_2 + \frac{1}{2} - sp}{\sigma}, \qquad \sigma = (spq)^{1/2}.$$

Example 2. In tossing six coins what is the probability of obtaining 2, 3, 4, or 5 heads?

Solution. We have $sp = 3$, $\sigma^2 = \frac{3}{2}$, $d_1 = 2$, $d_2 = 5$. Hence, $t_1 = -1.5/(\frac{3}{2})^{1/2} = -1.225$, $t_2 = 2.5/(\frac{3}{2})^{1/2} = 2.041$. Therefore,

$$Q = \int_{-1.225}^{2.041} = \int_0^{1.225} + \int_0^{2.041} = .38971 + .47932 = .869.$$

Although the use of Theorem IX assumes s large we obtain here with s small a good approximation to the exact value $Q = \frac{7}{8} = .875$. In this example it would have been a simple matter to evaluate and sum the terms of the binomial but when s is large and the range from d_1 to d_2 includes many terms this procedure may be very laborious. When this is the case the above theorem gives an approximation which may be quite satisfactory. The approximation is good if d_1 lies on one side of the mean and d_2 on the other at approximately equal distances.

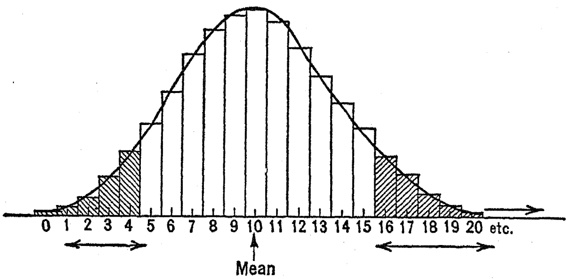


FIG. 3. BINOMIAL $(.8 + .2)^{50}$

Example 3. Suppose $p = .2$ is the probability of success in a single trial. Estimate the probability of obtaining less than five or more than fifteen successes in fifty trials.

Solution. The required probability, indicated by the shaded area in Figure 3, is $P = 1 - Q$ where Q is the probability of obtaining more than 4 and less than 16 successes. In using Theorem IX, we have

$$sp = 10, \qquad \sigma = 2.828, \qquad t_1 = -1.944, \qquad t_2 = 1.944.$$

Therefore,

$$P = 1 - Q = 1 - 2\int_0^{1.944} = .0519.$$

The exact probability is obtained by evaluating and adding the sixth to the sixteenth terms of $(.8 + .2)^{50}$ and subtracting the result from unity. However,

instead of computing these terms separately, a systematic procedure may be set up by which each term is made to depend upon the preceding term. Thus we may write a binomial as follows:

$$(q + p)^s = q^s(1 + k)^s = q^s\left\{1 + sk + \frac{s(s-1)}{2\,!}k^2 + \frac{s(s-1)(s-2)}{3\,!}k^3 + \cdots + k^s\right\}$$

where $k = \frac{p}{q}$. Then q^s may be computed by logarithms and its product with the terms in the brackets may be obtained on computing machines by a continuous process. Thus for the terms within the brackets,

the second term is first term multiplied by sk,

the third term is second term multiplied by $\frac{s-1}{2}k$,

the fourth term is third term multiplied by $\frac{s-2}{3}k$,

. . . .

the rth term is $(r - 1)st$ term multiplied by $\frac{s-(r-2)}{r-1}k$

In this way we find $Q = .9497$, so the required probability is $P = .0503$. For most practical purposes the approximation by use of the Theorem IX would be satisfactory.

Example 4. Find the probability that in throwing 100 coins one will obtain a number of heads which will differ from the expected number by less than five.

Solution.

$$t_1 = -\frac{4.5}{5} = -.9,$$

$$t_2 = \frac{4.5}{5} = .9.$$

So the required probability is given by

$$Q = 2\int_0^{.9} = .632.$$

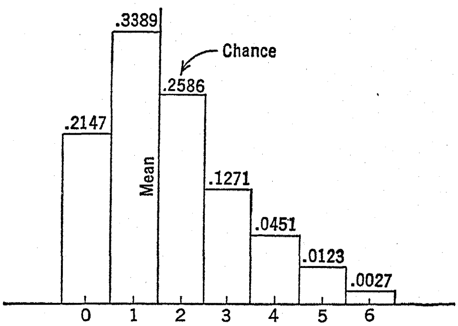


FIG. 4. FIRST SEVEN TERMS OF $(.95 + .05)^{30}$

Example 5. In the binomial $(.95 + .05)^{30}$ where $p = .05$ is the probability of success in a single trial, find the probability of as many as seven successes.

Solution. This binomial is too skew for a good fit with the normal curve, so the first seven terms of the expansion are evaluated. (See Figure 4.) Their sum is .9994 and this is the probability for less than seven successes. Therefore the probability for seven or more successes is .0006.

15. Simple Sampling of Attributes. It is a matter of common experience that certain fluctuations between observation and expectation under a given hypothesis may be explained on the basis of chance. For example, in throwing 100 coins an observed result of 45 heads and 55 tails does not warrant the conclusion that the coins are biased. In such cases a very natural question arises as to what sampling deviations may be allowed before we conclude that they indicate the operation of definite and assignable causes, *i.e.*, that the results are inconsistent with the given hypothesis. The theory dealing with such fluctuations in relative frequencies is called sampling of attributes.

Suppose we are given a sample of s individuals of which x have a certain character or attribute. The question then arises: Is this result consistent with the hypothesis that the sample is drawn from a population having the fraction p with the given character? Could it reasonably have arisen on the basis of chance or is it significant of other than chance factors? In answering this question our common-sense judgment is greatly aided by a probability scale for chance fluctuations under the given hypothesis. We therefore restate our question * more precisely as follows:

Suppose the probability of an event is known from theoretical considerations to be equal to p. What is the probability that in s trials the number of successes will differ numerically from the expected number $x = sp$ by as much as (or more than) an observed amount d?

The required probability may be estimated by means of the following corollaries to Theorem IX.

Corollary 1. *The probability that the number of successes x in s trials will differ from the expected number $x = sp$ by more than $|d|$ is approximately given by $P_\delta = 1 - Q_\delta$ where*

$$Q_\delta = 2\int_0^\delta \phi(t)dt \qquad \text{and} \qquad \delta = \frac{d + \frac{1}{2}}{\sigma}.$$

Corollary 2. *If the words "more than" in Corollary 1 be replaced by "as much as," then* $\delta = \dfrac{d - \frac{1}{2}}{\sigma}$.

The proofs are obvious if we admit that the normal curve fits the histogram of the point binomial.

* See *Problems in Sampling* — Camp, Journal American Statistical Association, p. 964, December, 1923.

In another slightly different form involving relative frequencies, Q_δ gives approximations to the probability that the difference between an observed relative frequency of success x/s and the true probability p satisfies the relation

$$\left|\frac{x}{s} - p\right| \leqq \delta \left(\frac{pq}{s}\right)^{1/2}$$

for every assigned positive value of δ.

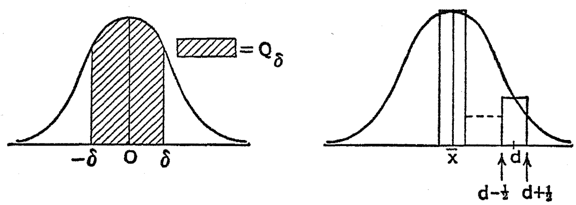


Fig. 5

In using Corollary 1, Table 2 gives a general idea of the magnitudes of probabilities for certain deviations. It is divided into two sections: the first section lists probabilities for specially selected deviations, the second section lists deviations for specially selected probabilities.

A computed probability is used to scale our judgment as to whether the deviation in question can be explained on the basis of chance.

Table 2. Abridged Normal Probability Scale

Deviation δ	*Chance of Deviation Outside* $\pm\,\delta$	*Deviation* δ	*Chance of Deviation Outside* $\pm\,\delta$
0.5	.617	.67	.50
1.0	.317	1.28	.20
1.5	.134	1.64	.10
2.0	.064	1.96	.05
2.5	.0124	2.33	.02
3.0	.0027	2.58	.01
3.5	.00047	2.88	.004

If it cannot be so explained, it is said to be "significant" of other than chance causes. In passing judgment on a deviation it is some-

times difficult to give a definite answer. Good judgment in these matters only comes from much experience in the particular field. However, we shall not often be wrong if we draw the following conventionalized conclusions about P_δ for a deviation outside $\pm\delta$:

If $P_\delta \geq .05$, δ is not significant,

If $P_\delta \leq .01$, δ is significant,

If $.05 > P_\delta > .01$, our conclusion about δ is doubtful and we cannot say with much certainty whether the deviation is significant or not until we have more information.

We see from Table 2 that this rule allows chance fluctuations to explain a deviation from the expected value of as much as 2.58 in standard units. In some situations it may be desirable to extend this range and place the bounds of chance fluctuations at $\delta = \pm 3$. There is then a correspondingly greater degree of certainty that deviations outside these limits are significant.

Example 6. (*Rietz*) A group of scientific men reported 1705 sons and 1527 daughters. The examination of these numbers brings up the following fundamental questions of simple sampling. Do these data conform to the hypothesis that $\frac{1}{2}$ is the probability that a child to be born will be a boy? That is, can the deviations be reasonably regarded as fluctuations in simple sampling under this hypothesis? In another form, what is the probability in throwing 3232 coins that the number of heads will differ from $(3232/2) = 1616$ by as much as $d = 1705 - 1616 = 89$?

Solution. $s = 3232$, $(pqs)^{1/2} = 28.425$, $\delta = \dfrac{88.5}{28.425} = 3.113$, $P_\delta = 1 - 2\displaystyle\int_0^{3.113}$ $= 1 - .9981 = .0019$.

Hence we conclude that these data cannot be explained on the basis of chance, *i.e.*, they are inconsistent with an hypothetical sex ratio of $\frac{1}{2}$.

16. Probable Error. The word *error* is technically used in statistics to denote a deviation from the expected value. The deviation δ for which $P_\delta = .5$ is commonly called "probable error." This term is misleading because it is not the most probable error. *Equally likely deviation* would be a more appropriate name for it.

From the normal probability scale we find that this deviation is $\delta = .6745$ in standard units or $.6745\sigma$ in arbitrary units. Hence for a normal distribution, probable error is equivalent to the quartile deviation which, in Part I, we have called E in x units and s in standard units. In other words, the probability is one-half that a

variate chosen at random will have a value within the range $E(x) \pm .6745\sigma_x$. This definition of probable error combines the assumption of a normal distribution with the specification of an even wager.

Used as a scale unit along the x-axis, probable error is sometimes simply defined as a yardstick which is approximately $\frac{2}{3}\sigma_x$. This definition does not impose the condition that the distribution necessarily follow the normal curve. But there is no real gain in the removal of this condition if, for an interpretation of the signficance of such a deviation, we must refer to a normal probability scale. That is, in testing the significance of a discrepancy between an observed value and the expected value there is no merit in expressing that discrepancy in multiples of approximately $\frac{2}{3}\sigma$ instead of σ itself. It would seem that the language of probable error should be abandoned.

17. Standard Error and Correlation of Errors in Class Frequencies. When the probability distribution of a variable is known the expected frequency in any class interval may be determined. Suppose we have obtained from a random sample of an infinite distribution an observed frequency distribution. The variates, N in number, should be distributed into n class intervals containing $f'_1, f'_2, \cdots f'_n$ each. Instead of this suppose we find $f_1, f_2, \cdots f_n$ where

$$\sum_1^n f'_i = N = \sum_1^n f_i.$$

Let Table 3 represent the two distributions.

Suppose next that a large number of such samples of N variates each are obtained under the same essential conditions. The ob-

TABLE 3

Class	*Class Mark*	*Observed Frequency*	*Theoretical Frequency*
1	x_1	f_1	f'_1
2	x_2	f_2	f'_2
·	·	·	·
·	·	·	·
i	x_i	f_i	f'_i
·	·	·	·
·	·	·	·
n	x_n	f_n	f'_n

served and expected distributions will not agree in practice unless the samples are distributed exactly as the universe from which they are drawn. In the above table, the x's are to be regarded simply as compartments and do not change. Only the frequencies change from sample to sample. Any class frequency f_s will vary from sample to sample, and these values of f_s will form a frequency distribution.

It is important in certain problems to have an expression for the expected value of the variance $\sigma_{f_s}{}^2$ of this distribution in terms of observed values. To derive this expression we let $p_s = f'_s/N$ be the probability that a variate will fall in the class s and $q_s = 1 - p_s$ be the probability that it will fall elsewhere. Then, considering the N variates as observations or trials, the theoretical distribution of frequency in this class will be given by $(q_s + p_s)^N$ and the square of the standard deviation of f_s in the theoretical distribution is given by

$$\sigma_{f_s}{}^2 = Np_sq_s.$$

If we accept the observed relative frequency f_s/N as an approximation to p_s then we have

$$\sigma_{f_s}{}^2 = N\frac{f_s}{N}\left(1 - \frac{f_s}{N}\right)$$

which reduces to

$$\sigma_{f_s}{}^2 = f_s\left(1 - \frac{f_s}{N}\right) \tag{10}$$

as an approximate* value of the desired expression.

We will next consider the correlation between deviations from the expected values of the frequencies in any two classes, say the sth and tth. Let δf_s be a deviation from the expected value or theoretical mean of the sth class corresponding to a deviation δf_t from the expected value of the tth class. Since the total frequency is N, $N - f_s$ is the frequency which is distributed in classes other than the s class. If we obtain an excess δf_s in the s class then $-\delta f_s$ must be distributed among the other classes. If deviations from the expected values are due only to random sampling fluctuations it is

* When the sample is small, researches have shown that a better approximation can be obtained by multiplying the right side of (10) by $N/(N - 1)$. See Rietz, *Mathematical Statistics*, pp. 120–122.

reasonable to assume that $-\delta f_s$ is distributed among the other classes in proportion to their expected frequencies. Therefore, as the contribution from the f_t class, we have the proportion $f_t/(N - f_s)$ and the number $(-\delta f_s)f_t/(N - f_s)$.

If the mean value of δf_t equals $-\delta f_s\, f_t/(N - f_s)$, for δf_s assigned, then $-f_t/(N - f_s)$ must be the regression coefficient of δf_t on δf_s. Therefore,

$$-\frac{f_t}{N - f_s} = r_{\delta f_t \delta f_s}\frac{\sigma_{\delta f_t}}{\sigma_{\delta f_s}} = r_{f_t f_s}\frac{\sigma_{f_t}}{\sigma_{f_s}}$$

so that

$$\begin{aligned} r_{f_t f_s}\sigma_{f_t}\sigma_{f_s} &= -\frac{f_t}{N - f_s}\sigma_{f_s}{}^2 \\ &= -\frac{f_t}{N(1 - p_s)}Np_s(1 - p_s) \\ &= -f_t p_s = -\frac{f_t f_s}{N}. \end{aligned}$$

Hence we have the result

$$r_{f_t f_s} = -\frac{\dfrac{f_t f_s}{N}}{\sigma_{f_t}\sigma_{f_s}}. \tag{11}$$

Clearly, $r_{\delta f_t \delta f_s} = r_{f_t f_s}$, and $\sigma_{\delta f_t}{}^2 = \sigma_{f_t}{}^2$, $\sigma_{\delta f_s}{}^2 = \sigma_{f_s}{}^2$, since the δ's measure deviations from their expected frequencies.*

For an application of the above formula and the Bernoulli Theory in general see *The Use of Statistical Techniques in Certain Problems of Market Research* — Brown. Publication of the Graduate School of Business Administration, Harvard University, vol. XXII, no. 3, 1935.

18. The Poisson Exponential. If p (or q) is small the normal curve cannot ordinarily be used with confidence to approximate the

* The correlation of errors here is properly a multivariate problem depending on the multinomial distribution. The argument given above indicates the plausibility of the result but it is not to be construed as a rigorous proof. By means of more advanced mathematics the correlation coefficient can be proved to have the result found without making use of the assumption that any excess frequency in one class is distributed among the other classes in proportion to their frequencies. In other words, the assumption is superfluous.

terms of the binomial $(q + p)^s$. If s is large but sp is in the neighborhood where x is small, a useful approximation to

$$(12) \qquad f(x) = \frac{s\,!}{x\,!\,(s - x)\,!}\,p^x q^{s-x}$$

may be given by means of the Poisson exponential function. Statistical examples of this situation are sometimes called rare events and occur in widely different fields; for example, the number born blind per year in a large city, the number of organisms of a given size S on a given glass slide that escape death by X-rays after being exposed for t seconds, the number of times in a certain year that the volume of trading on the New York Stock Exchange exceeds M million shares, the frequency of certain "peaks" in a given time interval such as occur in telephone "traffic," and other problems in demands for services.

Suppose, then, that p is the probability for the occurrence of the rare event in question and assume that $q = 1 - p$ is nearly unity. Let s be so large that $s\,!$ and $(s - x)\,!$ may be replaced by their Stirling approximations [cf. (12) of Chapter II]. Making these replacements, (12) becomes

$$(13) \qquad f(x) = \frac{s^{s+1/2}e^{-s}p^x q^{s-x}}{x\,!\,(s - x)^{s-x+1/2}e^{-s+x}}.$$

Writing the second factor in the denominator of (13) in the form $s^{s-x+1/2}(1 - x/s)^{s-x+1/2}$, it is readily seen that (13) becomes

$$f(x) = \frac{(sp)^x e^{-x}(1 - p)^{s-x}}{x\,!\left(1 - \dfrac{x}{s}\right)^{s-x-1/2}}.$$

Now when x is small and s is large,*

$$\left(1 - \frac{x}{s}\right)^{s-x+1/2} \approx \left(1 - \frac{x}{s}\right)^s$$
$$\approx e^{-x}$$

and

$$(1 - p)^{s-x} \approx (1 - p)^s$$
$$\approx e^{-sp}.$$

* The symbol $\approx$ is used to mean "approximately equal."

For the required approximation we have, therefore,

$$f(x) = \frac{m^x e^{-m}}{x\,!} \tag{14}$$

where $m = sp$. This is *Poisson's exponential function.* It is tabulated for various values of m and x in *Tables for Statisticians and Biometricians.* The terms of the series

$$e^{-m}\left(1 + m + \frac{m^2}{2\,!} + \frac{m^3}{3\,!} + \cdots + \frac{m^x}{x!}\right) \tag{15}$$

give the probability of exactly 0, 1, 2, $\cdots$, or x occurrences of the rare event in s trials. It is worthy of note that the Poisson exponential has only one parameter, m, whereas the normal curve has two parameters, the mean and σ.

Certain simple and interesting results may be obtained for the moments of the distribution given by (14) when x takes all integral values from $x = 0$ to $x = s$. First we observe that when $x = s$ in (15) we have

$$\sum_{x=0}^{s} \frac{m^x e^{-m}}{x\,!} = 1$$

approximately if s is large. Then

$$\begin{aligned} E(x) = \nu_1 &= \sum_{x=0}^{s} x f(x) \\ &= me^{-m}\left[1 + m + \frac{m^2}{2\,!} + \cdots + \frac{m^{s-1}}{(s-1)\,!}\right] \\ &\approx me^{-m}e^{m} \\ &= m = sp, \text{ approximately.} \end{aligned}$$

And

$$\begin{aligned} \nu_2 &= \sum_{0}^{s} x^2 f(x) \\ &= \sum_{0}^{s} [x(x-1) + x] f(x) \\ &= m(m+1), \text{ approximately.} \end{aligned}$$

From these results, we have

$$\text{Mean} = m = sp$$
$$\mu_2 = m(m + 1) - m^2$$
$$= m.$$
$$\therefore \quad \sigma = (m)^{1/2}.$$

It may also be shown that

$$\nu_3 = m(m^2 + 3m + 1)$$
$$\nu_4 = m(m^3 + 6m^2 + 7m + 1)$$

whence we find that

$$\mu_3 = m$$
$$\mu_4 = 3m^2 + m$$

and

$$\alpha_3 = \frac{1}{m^{1/2}}, \qquad \alpha_4 = 3 + \frac{1}{m}.$$

It is a rather striking result that each of the mean, variance, and μ_3 is equal to m.

The importance of the Poisson approximation in dealing with certain problems in telephone engineering and other fields is discussed in Fry's book, *Probability and Its Engineering Uses.* The interested student might investigate and prepare a special report on some of these applications.

Problems

1. Use Theorem IX to approximate the following sums:
   (*a*) the terms of $(\frac{1}{3} + \frac{2}{3})^{90}$ in which $50 \leq x \leq 70$.
   (*b*) the terms of $(.946 + .054)^{521}$ in which $x \geq 34$.
2. Fit a normal curve to the point binomial $(\frac{1}{2} + \frac{1}{2})^4$.
3. Fit a normal curve to $(\frac{1}{2} + \frac{1}{2})^6$.
4. Suppose you are studying IQ's and it is known that 20% in the universe with which you are dealing have an IQ below M, so that $\frac{1}{5}$ is the probability that an individual chosen at random has an IQ below M. (M itself has no bearing on the solution of the problem.) If a teacher had a class of fifty which could be regarded as a random sample from this universe, would it be exceptional if she found less than five or more than fifteen with IQ's below M? (See Example 3.)
5. Vital statistics gathered over a long period of time indicate that 5% of patients suffering from a certain disease die from that disease. Suppose that out of 30 cases examined in a certain city seven deaths were reported. Was this unusual? (See Example 5.)

6. (*Camp*) A dean's report showed the following figures:

Subject	*Honor Grades*		*Failures*		*Number Examined*
	Number	%	*Number*	%	
German	187	36	33	6.3	521
Mathematics	162	35	38	8.2	466
Music	11	50	0	0.0	22
All Subjects		38		5.4	

Taking $p = .38$ for honor grades and $p = .054$ for failures find the probability: (*a*) that in selecting at random (from a supposedly infinite number), one would obtain as few honor grades as were obtained in German; (*b*) as many failures; (*c*) in selecting 466 at random, one would obtain as few honor grades as were obtained in mathematics; (*d*) as many failures; (*e*) in selecting 22 at random, one would obtain no failures (as in music); (*f*) eleven or more honor grades.

Hints. (*a*) Find sum of terms of $(.62 + .38)^{521}$ in which $x \leq 187$.

(*b*) See Problem 1 (*b*) above.

(*e*) Evaluate $(0.54)^{22}$ by logarithms.

7. (*Burgess*) If analyzed past experience shows that 4% of all insured white males of exact age 65 have died within a year, and it is found that 60 of a similar group of 1000 actually die within a year, should the group be regarded as essentially different from the general mass — that is, is the departure from the expected mortality greater than might be expected as a result of chance variation alone?

8. (*Richardson*) In a coin tossing experiment in which a coin was tossed 400 times, 250 heads appear. Do you believe the experiment was honestly performed?

9. (*Lovitt* and *Holtzclaw*) Would you be willing to bet 10 to 1 that an opponent could not throw the sum 7 with two dice at least 23 times in a hundred throws with two dice?

10. (*Lovitt* and *Holtzclaw*) The 1919 report of the Census Bureau in its bulletin on *Mortality Statistics* shows the average death rate from tuberculosis (all forms) for the period 1906–1910 to be 163.5 per 100,000 of population and $\sigma = 12.78$.

In the following instances is the variation from the average such as to justify one in constructing a theory as to the causes of this variation?

California	210.4
Colorado	244.2
Michigan	99.7
N. Y. Bronx	445.7
Scranton, Pa.	97.4.

11. A sociologist who is interested in the characteristics of a certain race which we will call R, hit on the idea of trying to sort R's from non-R's in the writings of unknown persons. Accordingly he persuaded a colleague to let him have 64 examination papers, with names removed, from psychology classes at Blank University. On 43 of these papers he correctly spotted the students as R's or non-R's. In 21 cases he missed. Find the probability of this performance having resulted from pure chance.

12. A coin is tossed s times. It is desired that the relative frequency of the appearance of heads shall not be greater than .51 or less than .49. Find the smallest value of s that will insure the above results with a degree of certainty $Q_\delta \geq .90$.

Solution. We must determine s such that $Q_\delta = .90$ (at least) that

$$\left|\frac{x}{s} - \frac{1}{2}\right| \leq .01.$$

We have

$$\delta\left(\frac{pq}{s}\right)^{1/2} = .01$$

$$\delta = .02\sqrt{s}$$

since $p = q = \frac{1}{2}$. Also

$$Q_\delta = 2\int_0^\delta \phi(t)\,dt = .90$$

whence from the tables, we find $\delta = 1.645$. Therefore,

$$.02\sqrt{s} = 1.645$$

and

$$s = 6745.$$

13. A coin is tossed s times. It is desired that the relative frequency of the appearance of heads shall not be greater than .502 or less than .498. Find the smallest value of s that will insure the foregoing results with a degree of certainty $Q_\delta \geq \frac{15}{16}$.

14. (*Camp*) A census report showed that in general 59.58% of New York City children went to school, but that only 56.8% of the negro children went to school. The number of negro children was 20,000. Was the difference due to chance?

15. Read and give a report on the reference given at the end of § 17.

16. Find applications of the Poisson exponential function in the literature and report on them in class.

CHAPTER II

SOME USEFUL INTEGRALS AND FUNCTIONS

To avoid interruption later on we will discuss here certain integrals and functions which will be useful in subsequent chapters.

1. The Gamma Function. The improper integral

$$(1) \qquad \Gamma(n) = \int_0^\infty x^{n-1}e^{-x}\,dx, \qquad n > 0,$$

is called the Gamma function of the positive number n. The difference equation

$$(2) \qquad \Gamma(n+1) = n\Gamma(n)$$

is easily established from (1) by integration by parts (see the chapter on the Gamma function in any textbook on advanced calculus). By successive reduction of (2) we obtain

$$\Gamma(n+1) = n(n-1)\cdots(n-k)\Gamma(n-k)$$

where k is a positive integer less than n. If n is also a positive integer and $k = n - 1$ then we have

$$(3) \qquad \Gamma(n+1) = n\,!$$

since from (1), $\Gamma(1) = 1$. Because of (3) the Gamma function is sometimes called the factorial function. It may be considered as a generalization of $n\,!$ when n is fractional. The graph of the function defined in (1) is shown in Figure 6. It can be drawn from the following values, some of which follow immediately from (2) and the others will be established later.

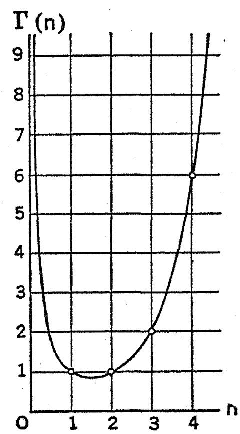


FIG. 6

$\Gamma(0) = \infty$	$\Gamma(2) = 1.$
$\Gamma(1) = 1$	$\Gamma(3) = 2.$
$\Gamma(\frac{1}{2}) = (\pi)^{1/2}$	$\Gamma(4) = 6.$

Other forms of (1) may be obtained by changes of variable. For example,

$$\Gamma(n) = 2\int_0^\infty y^{2n-1}e^{-y^2}\,dy, \qquad \text{by } x = y^2. \tag{4}$$

From this form we can show that

$$\int_0^\infty e^{-y^2}\,dy = \tfrac{1}{2}(\pi)^{1/2}. \tag{5}$$

To establish (5) we first observe from (4) that

$$\Gamma(\tfrac{1}{2}) = 2\int_0^\infty e^{-y^2}\,dy. \tag{6}$$

Since (6) is independent of the variable of integration, we may also write

$$\Gamma(\tfrac{1}{2}) = 2\int_0^\infty e^{-x^2}\,dx.$$

So

$$[\Gamma(\tfrac{1}{2})]^2 = 4\int_0^\infty e^{-x^2}\,dx \int_0^\infty e^{-y^2}\,dy$$

$$= 4\int_0^\infty \int_0^\infty e^{-(x^2+y^2)}\,dx\,dy, \tag{7}$$

the passage from the product of two integrals to the double integral being valid since neither the limits nor the integrand of either integral depend on the variable in the other.

To evaluate (7) it will be convenient to change to polar coördinates. First, however, we will make a few remarks about a change of variables in general. Let x and y be the coördinates of a point with respect to a set of rectangular axes in a plane, u and v the coördinates of another point with respect to a similarly chosen set of rectangular axes in some other plane. Suppose we have a function of the variables (x, y),

$$z = f(x, y),$$

and we make x and y depend on new variables u and v by the relations

$$x = g(u, v) \qquad \text{and} \qquad y = h(u, v).$$

These relations establish a certain correspondence between the points of the two planes. Let dA be an element of area for the function $f(x, y)$. Then it is shown in advanced calculus * that

$$dA = \left| J\left(\frac{x, y}{u, v}\right) \right| du\, dv$$

where $\left|J\left(\frac{x, y}{u, v}\right)\right|$ is a convenient symbol for the absolute value of the determinant

$$\begin{vmatrix} \dfrac{\partial x}{\partial u} & \dfrac{\partial x}{\partial v} \\ \\ \dfrac{\partial y}{\partial u} & \dfrac{\partial y}{\partial v} \end{vmatrix} = \frac{\partial x}{\partial u}\frac{\partial y}{\partial v} - \frac{\partial x}{\partial v}\frac{\partial y}{\partial u}$$

and the latter is called the *Jacobian* or *functional determinant* of the transformation.

If, then, we change (7) to polar coördinates by letting

$$(8) \qquad \begin{cases} x = r \cos \theta \\ y = r \sin \theta \end{cases}$$

the Jacobian is

$$\begin{vmatrix} \cos \theta & -r \sin \theta \\ \sin \theta & r \cos \theta \end{vmatrix} = r.$$

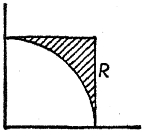
Therefore, the element of integration $dx\, dy$ becomes $r\, dr\, d\theta$. The limits of integration are now from 0 to ∞ for r and from 0 to $\pi/2$ for θ. From (8), $x^2 + y^2 = r^2$. So (7) becomes †

$$[\Gamma(\tfrac{1}{2})]^2 = 4 \int_0^{\pi/2} \int_0^{\infty} e^{-r^2}\, r\, dr\, d\theta$$

$$= 2 \int_0^{\pi/2} d\theta = \pi.$$

* See *Mathematical Analysis*, Goursat-Hedrick, vol. 1.

† The transformation to polar coördinates and subsequent integration involves a remainder term T which is the integral over an area between a quadrant of radius R and a square of side R. But it can be shown that $T \to 0$ as $R \to \infty$. (Cf. Wilson's *Advanced Calculus*, p. 364.)



Hence,

$$\Gamma(\tfrac{1}{2}) = (\pi)^{1/2}, \tag{9}$$

and from (9) and (6) we obtain (5).

For a more general form of (5) we may let $y = t/(2k)^{1/2}$, $k > 0$, and obtain

$$\int_0^\infty e^{-t^2/2k}\,dt = \tfrac{1}{2}(2\pi k)^{1/2}, \tag{10}$$

and

$$\int_{-\infty}^\infty e^{-t^2/2k}\,dt = (2\pi k)^{1/2}. \tag{10a}$$

An alternate derivation of (9) may be given as follows. The right-hand member of (7) represents the volume V under the bell-shaped surface

$$z = e^{-(x^2+y^2)} \tag{11}$$

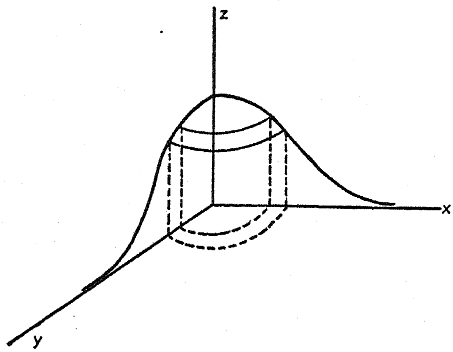


Fig. 7

and so from (7) we have $\Gamma(\tfrac{1}{2}) = V^{1/2}$. Since (11) is a surface of revolution we may take as the element of volume a cylindrical shell of radius r, thickness dr, and height z. Then

$$dV = 2\pi r\,dr\,z = 2\pi r e^{-r^2}\,dr,$$

$$V = 2\pi \int_0^\infty e^{-r^2} r\,dr = \pi,$$

and consequently we obtain (9).

2. Stirling's Approximation. An asymptotic expression, that is, an approximation with small percentage error, may be obtained for $n\,!$ when n is large. The following formula

$$n\,! = (2\pi)^{1/2} n^{n+1/2} e^{-n} \tag{12}$$

is called *Stirling's approximation.* A closer approximation is

$$n\,! = (2\pi)^{1/2} n^{n+1/2} e^{-n}\left(1 + \frac{1}{12n} + \cdots\right).$$

However, the first term usually gives sufficiently close approxima-

tions if n is fairly large. A derivation of (12) may be found in several places. Among these are

Probability and Its Engineering Uses — Fry, D. Van Nostrand Company; and *Introduction to Mathematical Probability* — Uspensky, McGraw-Hill Company. Seven-place tables of log n ! up to $n = 1000$ are given in *Glover's Tables*.

3. The Beta Function. The definite integral

$$\text{B}(m, n) = \int_0^1 x^{m-1}(1 - x)^{n-1}\, dx \tag{13}$$

is called the Beta function of any two positive numbers m and n. Another useful form is

$$\text{B}(m, n) = 2\int_0^{\pi/2} \sin^{2m-1}\theta \cos^{2n-1}\theta\, d\theta \tag{14}$$

which is obtained by letting $x = \sin^2\theta$ in (13).

If we let $x = 1 - y$, (13) becomes

$$\begin{aligned} \text{B}(m, n) &= \int_0^1 (1 - y)^{m-1}y^{n-1}\, dy \\ &= \int_0^1 (1 - x)^{m-1}x^{n-1}\, dx \\ &= \text{B}(n, m). \end{aligned}$$

Therefore, m and n may be interchanged.

A relation between the Beta and Gamma functions may be obtained as follows. From (4) we may write

$$\begin{aligned} \Gamma(n)\Gamma(m) &= 4\int_0^\infty x^{2n-1}e^{-x^2}\, dx \int_0^\infty y^{2m-1}e^{-y^2}\, dy \\ &= 4\int_0^\infty\int_0^\infty x^{2n-1}y^{2m-1}e^{-(x^2+y^2)}dx\, dy. \end{aligned}$$

Since the region of integration is the first quadrant of the xy-plane we have, upon changing to polar coördinates,

$$\begin{aligned} \Gamma(n)\Gamma(m) &= 4\int_0^{\pi/2}\int_0^\infty r^{2(m+n-1)}e^{-r^2} \sin^{2m-1}\theta \cos^{2n-1}\theta r\, d\theta\, dr \\ &= 4\int_0^{\pi/2} \sin^{2m-1}\theta \cos^{2n-1}\theta\, d\theta \int_0^\infty r^{2(m+n)-1}e^{-r^2}\, dr \\ &= \text{B}(m, n)\Gamma(m + n), \end{aligned}$$

by (14) and (4). Hence

$$\text{(15)} \qquad B(m, n) = \frac{\Gamma(m)\Gamma(n)}{\Gamma(m+n)}.$$

4. Reduction to Gamma and Beta Functions. By appropriate changes of variables many of the integrals that occur in statistics may be evaluated by expressing them in terms of Gamma and Beta functions.

Examples

(*a*) Prove that

$$\int_0^\infty y^{N-1}e^{-Ny^2/2\sigma^2}\,dy = \frac{1}{2}\left(\frac{2\sigma^2}{N}\right)^{N/2}\Gamma\left(\frac{N}{2}\right)$$

Solution. This integral may be written

$$\tfrac{1}{2}\int_0^\infty (y^2)^{(N-2)/2}e^{-Ny^2/2\sigma^2}\,d(y^2).$$

By the substitution

$$x = \frac{Ny^2}{2\sigma^2}, \qquad d(y^2) = \frac{2\sigma^2}{N}\,dx$$

this becomes

$$\frac{1}{2}\left(\frac{2\sigma^2}{N}\right)^{N/2}\int_0^\infty x^{(N-2)/2}e^{-x}\,dx =$$

$$\frac{1}{2}\left(\frac{2\sigma^2}{N}\right)^{N/2}\Gamma\left(\frac{N}{2}\right).$$

(*b*) Determine k so that

$$k\int_0^\infty e^{-Ns^2/2\sigma^2}(s^2)^{(N-3)/2}\,d(s^2) = 1.$$

Solution. By the substitution

$$x = \frac{Ns^2}{2\sigma^2}, \qquad d(s^2) = \frac{2\sigma^2}{N}\,dx$$

this becomes

$$k\left(\frac{2\sigma^2}{N}\right)^{(N-1)/2}\int_0^\infty x^{(N-3)/2}e^{-x}\,dx = 1$$

and so

$$k = \frac{\left(\dfrac{N}{2\sigma^2}\right)^{(N-1)/2}}{\Gamma\left(\dfrac{N-1}{2}\right)}.$$

(c) Determine K so that $K\int_{-\infty}^{\infty}(1+z^2)^{-N/2}\,dz = 1$.

Solution. By the substitution $z = \tan\theta$ this becomes $2K\int_0^{\pi/2}\cos^m\theta\,d\theta$ where $m = N - 2$. From Exercise 9 below we find that

$$\frac{1}{K} = \mathrm{B}\left(\frac{m+1}{2}, \frac{1}{2}\right)$$

$$= \frac{\Gamma\left(\frac{m+1}{2}\right)\Gamma\left(\frac{1}{2}\right)}{\Gamma\left(\frac{m+2}{2}\right)}$$

whence

$$K = \frac{\Gamma\left(\frac{N}{2}\right)}{(\pi)^{1/2}\Gamma\left(\frac{N-1}{2}\right)}.$$

5. Incomplete Beta and Gamma Functions. The integral

$$\Gamma_x(n+1) = \int_0^x e^{-x}x^n\,dx \tag{16}$$

is called the incomplete Gamma function. Similarly

$$\mathrm{B}_x(m, n) = \int_0^x x^{m-1}(1-x)^{n-1}\,dx \tag{17}$$

is called the incomplete Beta function. Both (16) and (17) are useful functions in mathematical statistics and they have been tabulated by Karl Pearson and his staff at the Biometric Laboratory, University College, London. They are published by the Cambridge University Press.

Exercises

1. Show that the Gamma function becomes infinite when $n = 0$. *Hint.* From (2) you can obtain

$$\Gamma(n+k) = (n+k-1)\cdots(n+1)n\Gamma(n),$$

that is

$$\Gamma(n) = \frac{\Gamma(n+k)}{n(n+1)\cdots(n+k-1)}.$$

2. Show that $\int_{-\infty}^{\infty}\phi(t)\,dt = 1$ where $\phi(t) = \dfrac{1}{(2\pi)^{1/2}}e^{-t^2/2}$.

3. Prove that $\Gamma(\tfrac{1}{2}) = (\pi)^{1/2}$.

4. Evaluate $\int_{-2}^{\infty} e^{-3x}(1 + x/2)^7\,dx$ by transforming it into a Gamma function.
   Hint. Let $cy = 1 + x/2$ and determine c so that $e^{-3x} = ke^{-y}$.
   Ans. $(e^6 7\,!)/3(6^7)$.

5. Evaluate $\int_{6}^{\infty} e^{-2x}(x - 6)^7\,dx$. *Ans.* $e^{-12}2^{-8}7\,!$.

6. Evaluate $\int_{0}^{\infty} e^{-x}x^{1/2}\,dx$, given that $\int_{0}^{\infty} e^{-x}x^{-1/2}\,dx = (\pi)^{1/2}$.
   Hint. $\Gamma(\frac{3}{2}) = \frac{1}{2}\Gamma(\frac{1}{2})$.

7. Find the difference and the ratio between the exact value of 10 ! and the approximate value obtained by using Stirling's formula.

8. Using (15) show that

$$\frac{\Gamma\left(\frac{N}{2}\right)}{(\pi)^{1/2}\Gamma\left(\frac{N-1}{2}\right)} = \frac{1}{B[\frac{1}{2}(N-1),\ \frac{1}{2}]}.$$

9. Prove that $\int_{0}^{\pi/2} \cos^m \theta\,d\theta = \frac{1}{2}\,B[(m+1)/2,\ \frac{1}{2}]$. *Hint.* Use (14).

10. Given that $f(n) = n^{1/2}B(n/2,\ \frac{1}{2})$, show that $\lim_{n\to\infty} f(n) = (2\pi)^{1/2}$.

CHAPTER III

GENERAL CONCEPT OF DISTRIBUTION FUNCTION OF A CONTINUOUS VARIABLE. GENERALIZED FREQUENCY CURVES

1. Fundamental Notions and Definitions. The notion of distribution functions relates to theoretical universes. The concept is an idealization of observed distributions comparable to the idealization of the outlines of material objects into the straight lines and circles of geometry.

A continuous variable x is said to have the distribution function $f(x)$, which we take to be single-valued and non-negative, if the frequency of occurrence of x in the range $a < x < b$ is measured by

$$\int_a^b f(x)\,dx. \tag{1}$$

If x has the distribution function $f(x)$ with total frequency N, then

$$\int_{-\infty}^{\infty} f(x)\,dx = N, \tag{2}$$

and $y = f(x)$ is called a *theoretical frequency curve* or, more briefly, a *frequency curve.* If the actual occurrence of the variable is limited to a finite range, $f(x)$ is defined to be identically zero outside that range. If the total area under the curve is taken as unity, so that

$$\int_{-\infty}^{\infty} f(x)\,dx = 1, \tag{3}$$

then $y = f(x)$ is variously called the *probability density*, the *probability distribution*, or the *probability function* of x. Then, $f(x)\,dx$ gives, to within infinitesimals of order higher than that of dx, the probability that x lies in the interval $(x, x + dx)$. Under condition (3), the integral (1) denotes the *probability* that x lies in the interval (a, b). Under condition (2), (1) denotes the *frequency* of values in the interval (a, b). A distribution function can be regarded, therefore, either as a frequency curve or as a probability curve according as condition (2) or (3) is imposed. The distinction can be adjusted by determining appropriately a constant factor in $y = f(x)$.

2. Moments. If x is distributed in accord with the frequency curve $y = f(x)$, with total frequency N, the moment of order k about the y-axis is defined by

$$\nu_k = \frac{1}{N}\int_{-\infty}^{\infty} x^k f(x)\, dx. \tag{4}$$

In particular, for $k = 1$ we have the mean, $\nu_1 = \bar{x}$,

$$\bar{x} = \frac{1}{N}\int_{-\infty}^{\infty} x f(x)\, dx.$$

If the mean is taken as the origin of measurement, so that

$$\int_{-\infty}^{\infty} (x - \bar{x}) f(x)\, dx = 0,$$

then the moment of order k about the mean is defined by

$$\mu_k = \frac{1}{N}\int_{-\infty}^{\infty} (x - \bar{x})^k f(x)\, dx. \tag{5}$$

In particular, when $k = 2$ we have the variance, $\mu_2 = \sigma^2$,

$$\sigma^2 = \frac{1}{N}\int_{-\infty}^{\infty} (x - \bar{x})^2 f(x)\, dx.$$

The μ's can be expressed in terms of the ν's by the relation

$$\begin{aligned}\mu_k = {} & \nu_k - C(k, 1)\nu_{k-1}\nu_1 + C(k, 2)\nu_{k-2}\nu_1^2 - \cdots \\ & + (-1)^r C(k, r)\nu_{k-r}\nu_1^r + \cdots + (-1)^{k-1}[C(k, k-1) - 1]\nu_1^k\end{aligned} \tag{6}$$

where
$$C(k, r) = \frac{k\,!}{(k - r)\,!\, r\,!}.$$

In particular, the following relations are useful in computations:

$$\begin{cases} \mu_2 = \nu_2 - \nu_1^2 \\ \mu_3 = \nu_3 - 3\nu_2\nu_1 + 2\nu_1^3 \\ \mu_4 = \nu_4 - 4\nu_3\nu_1 + 6\nu_2\nu_1^2 - 3\nu_1^4. \end{cases} \tag{7}$$

The first of (7) is proved below and the others may be established in a similar way.

$$\begin{aligned}\mu_2 &= \frac{1}{N}\int_{-\infty}^{\infty} x^2 f(x)\, dx - \frac{2\bar{x}}{N}\int_{-\infty}^{\infty} x f(x)\, dx + \frac{\bar{x}^2}{N}\int_{-\infty}^{\infty} f(x)\, dx \\ &= \frac{1}{N}\int_{-\infty}^{\infty} x^2 f(x)\, dx - \bar{x}^2 = \nu_2 - \nu_1^2.\end{aligned}$$

In standard units the moment of order k is defined by

$$\alpha_k = \frac{\mu_k}{\sigma^k} \tag{8}$$
$$= \int_{-\infty}^{\infty} t^k h(t)\, dt,$$

where

$$\left\{\begin{aligned} t &= \frac{(x - \bar{x})}{\sigma} \\ h(t) &= \frac{\sigma}{N} f(\sigma t + \bar{x}) = \frac{\sigma}{N} f(x). \end{aligned}\right. \tag{9}$$

From (8) we have

$$\alpha_0 = 1,$$
$$\alpha_1 = 0,$$
$$\alpha_2 = 1.$$

Analogous definitions of moments could be given for probability functions. When $N = 1$, in accordance with (3), the integrals in (4) and (5) are also called *expected values.* The language of expected values will be used in another chapter where we will be dealing more with probability functions. Before proceeding with the discussion of frequency curves, however, we will give an example of a probability curve.

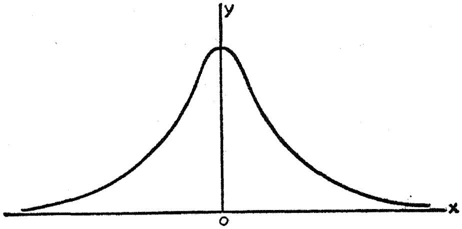


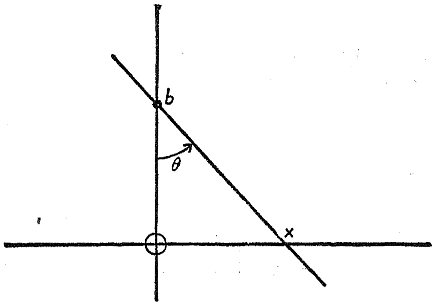
FIG. 8. THE CAUCHY CURVE

Example. The Cauchy curve is a classical example of a probability distribution although its use in present day statistics is relatively unimportant. Its equation is

$$y = \frac{b}{\pi(b^2 + x^2)}, \quad -\infty \leq x \leq \infty, \quad b > 0. \tag{10}$$

The curve is symmetrical having its center at $x = 0$.

A simple derivation of this function is as follows. For a given real constant b locate the point $(0, b)$ as in the figure below. Let lines be drawn at random through $(0, b)$ and let θ be the variable angle between any such line and the negative direction of the y-axis; θ varies between the limits $-\pi/2$ and $\pi/2$. The hypothesis is that all values of θ in this range are equally likely. Denote the intercepts on the horizontal axis by x. Clearly, $-\infty < x < \infty$. The relation between θ and x is



$$\theta = \tan^{-1}\frac{x}{b}.$$

Under the hypothesis, the probability that an angle $0bx$ will be contained between θ and $\theta + d\theta$ is $d\theta/\pi$. By differentiation we find the relation between $d\theta$ and dx to be

$$\frac{d\theta}{\pi} = \frac{b\,dx}{\pi(b^2 + x^2)}. \tag{11}$$

Therefore, the points of intersection of the lines with the x-axis are distributed so that the probability that a value of x will fall in the range dx is given by the right-hand member of (11). Hence the probability function for the variable x is

$$f(x) = \frac{b}{\pi(b^2 + x^2)}$$

and the probability that x lies in a finite interval (c, d) is given by

$$P(c, d) = \int_c^d \frac{b\,dx}{\pi(b^2 + x^2)},$$

since the integral of the right-hand member of (11) from $-\infty$ to ∞ is equal to unity as can easily be verified. However, we cannot speak here of the mean value of x or of moments of higher order, since the integral

$$\int_{-\infty}^{\infty} \frac{x^k\,dx}{(b^2 + x^2)}$$

has no meaning for $k \geq 0$. This restriction does not apply to probability functions in general.

3. The Pearson System. There are two systems of generalized frequency curves in common use: the *Pearson system* and the *Gram-Charlier system.*

During the years 1895–1916 Karl Pearson published papers in which he showed that a set of frequency curves could be obtained by assigning values to the parameters in a certain first order differential equation. The Pearson school claims that all the different

types of frequency distributions that arise in practical statistics can be represented by the solutions of this equation.

With regard to the genesis of the Pearson system, one point of view is to regard it as empirical. Thus, starting with the differential equation

$$\frac{dy}{dt} = \frac{(m - t)y}{a + bt + ct^2}, \tag{12}$$

it is observed that the solutions of (12) must satisfy certain geometrical properties of unimodal frequency distributions, namely, (a) the curve should vanish at the ends of the range, *i.e.*, as $y \to 0$, $dy/dt \to 0$; (b) when $t = m$, corresponding to a mode, $dy/dt = 0$.

Among the solutions* of (12) there are several types of curves, the shapes depending on the parameters a, b, c, and m. Examples of symmetrical, skewed, U-shaped and J-shaped curves with finite and infinite range in either or both directions, are shown in Figure 9.

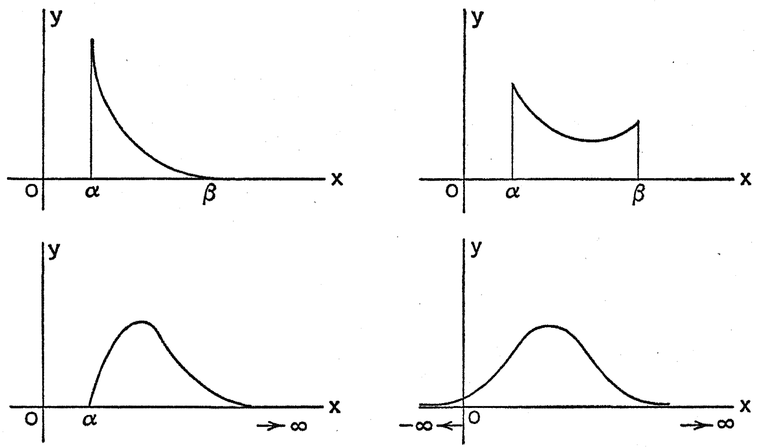


FIG. 9. TYPICAL CURVES OF THE PEARSON SYSTEM

The parameters in (12) can be expressed in terms of the moments of the system. Multiplying (12) by $t^k\,dt$ and integrating over all admissible values of t, we have

$$\int \frac{dy}{dt}(at^k + bt^{k+1} + ct^{k+2})\,dt = \int y(mt^k - t^{k+1})\,dt. \tag{13}$$

* What we actually do is to derive equations under the stated assumption that $y \to 0$ as $dy/dt \to 0$, and then generalize the results so as to admit as distribution functions solutions which do not vanish at the end(s) of the range.

Integrating the left-hand side by parts, we obtain

$$(14)\quad t^k(a + bt + ct^2)y] - \int y[akt^{k-1} + b(k+1)t^k + c(k+2)t^{k+1}]\,dt$$
$$= \int y(mt^k - t^{k+1})\,dt.$$

If yt^{k+2} vanishes at the ends of the range, then the first expression in (14) vanishes. If, in (12), $y = h(t)$ we have from (8) and (14),

$$(15)\qquad m\alpha_k + ak\alpha_{k-1} + b(k+1)\alpha_k + c(k+2)\alpha_{k+1} = \alpha_{k+1}.$$

Assigning k successively the values $k = 0, 1, 2, 3$, we obtain from (15) the four equations

$$(16)\qquad \begin{cases} m + b = 0 \\ a + 3c = 1 \\ m + 3b + 4c\alpha_3 = \alpha_3 \\ m\alpha_3 + 3a + 4b\alpha_3 + 5c\alpha_4 = \alpha_4 \end{cases}$$

from which the parameters can be determined. Solving (16) we obtain

$$(17)\qquad \begin{cases} m = \dfrac{1}{D}[\alpha_3\,(3 + \alpha_4)] \\[2mm] a = \dfrac{1}{D}[3\alpha_3^2 - 4\alpha_4] \\[2mm] b = \dfrac{1}{D}[-\alpha_3(3 + \alpha_4)] \\[2mm] c = \dfrac{1}{D}[6 + 3\alpha_3^2 - 2\alpha_4] \\[2mm] D = 18 + 12\alpha_3^2 - 10\alpha_4. \end{cases}$$

Carver* has expressed (17) in the more convenient form

$$(18)\qquad \begin{cases} m = -\dfrac{\alpha_3}{2(1 + 2\delta)}, \quad b = \dfrac{\alpha_3}{2(1 + 2\delta)}, \\[2mm] a = \dfrac{2 + \delta}{2(1 + 2\delta)}, \quad c = \dfrac{\delta}{2(1 + 2\delta)}, \\[2mm] \delta = \dfrac{2\alpha_4 - 3\alpha_3^2 - 6}{\alpha_4 + 3}. \end{cases}$$

* See the *Handbook of Mathematical Statistics* — Rietz *et al.*

Substitution of the above values into (15) yields an important recursion formula for the moments of the Pearson system:

$$\alpha_{k+1} = \frac{k}{2 - (k-2)\delta}[(2+\delta)\alpha_{k-1} + \alpha_3\alpha_k]. \tag{19}$$

For our purposes the most important curves in the Pearson system are the Type VII (normal curve) and Type III. These will now be discussed in some detail.

Type VII. If $\alpha_3 = 0 = \delta$, then (12) becomes

$$\frac{dy}{y} = -t\,dt$$

which upon integration yields the so-called normal curve

$$y = Ce^{-t^2/2}, \qquad -\infty < t < \infty. \tag{20}$$

The constant C may be determined so that the area under the curve is N. Imposing this condition and making use of (10*a*) of Chapter II we find that $C = N/(2\pi)^{1/2}$, and so (20) becomes

$$y = \frac{N}{(2\pi)^{1/2}}e^{-t^2/2}.$$

It is conventional to write this in the form

$$y = \frac{N}{\sigma}\phi(t) \tag{21}$$

where

$$\phi(t) = \frac{1}{(2\pi)^{1/2}}e^{-t^2/2}$$

$$t = \frac{(x - \bar{x})}{\sigma}.$$

We may call $\phi(t)$ the normalized normal curve.

Type III. If $\delta = 0$ but $\alpha_3 \neq 0$ we see from (18) that (12) becomes

$$\frac{dy}{dt} = \frac{-\left(\frac{\alpha_3}{2} + t\right)y}{1 + \frac{\alpha_3}{2}t}$$

which upon integration yields the Type III curve

$$y = K(A + t)^{A^2-1}e^{-At} \tag{22}$$

where $A = 2/\alpha_3$, the range being $(-A, \infty)$. The criterion for a Type III curve is that $\delta = 0$. That is, if a Type III curve is to represent an observed distribution the observed moments should satisfy, at least approximately, the relation

$$2\alpha_4 - 3\alpha_3^2 - 6 = 0.$$

Definitions of moments of an observed distribution are given in Part I.

The constant K in (22) may be determined by the condition

$$\int_{-A}^{\infty} y\, dt = N. \tag{23}$$

This integral can be evaluated by means of the Gamma function. Let $A^2 = n/2$ and let $A(A + t) = \chi^2/2$. Then we have $(A + t)^{A^2-1} = (\chi^2/2)^{(n/2)-1}A^{1-(n/2)}$, $e^{-At} = e^{n/2}e^{-\chi^2/2}$, and $dt = d(\chi^2)/2A$. Making these substitutions in (23) we obtain

$$KA^{-n/2}e^{n/2}\int_0^{\infty}\left(\frac{\chi^2}{2}\right)^{(n/2)-1}e^{-\chi^2/2}d\left(\frac{\chi^2}{2}\right) = N$$

and therefore

$$K = \frac{NA^{n/2}e^{-n/2}}{\Gamma\left(\frac{n}{2}\right)}.$$

So with χ^2 as the independent variable, (22) becomes

$$y = \frac{N}{2\Gamma\left(\frac{n}{2}\right)}\left(\frac{\chi^2}{2}\right)^{(n/2)-1}e^{-\chi^2/2}. \tag{22a}$$

When $N = 1$, (22*a*) defines the *probability distribution of* χ^2. This is an important function which we shall use in subsequent discussions.

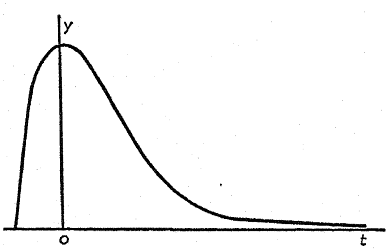


FIG. 10. TYPE III CURVE WHEN $|\alpha_3| < 2$

The designation "Type III" is usually restricted to the case for which $A^2 \neq 1$. When $A^2 > 1$, that is, when $|\alpha_3| < 2$, the curve is bell-shaped as shown in Figure 10.

In the Pearson system, the distance between the mean and mode is $m = -\alpha_3/2(1 + 2\delta)$, and is a measure of skewness.

Under the conditions imposed for Type VII, $m = 0$. For Type III, however, $m = -\alpha_3/2$ and therefore we have

$$|\text{mean} - \text{mode}| = \frac{\alpha_3}{2}.$$

Because of this relation, $|\alpha_3|/2$ is sometimes used as a measure of skewness in observed distributions. The curve for $\alpha_3 = -k$ (k = a constant) is a reflection of that for $\alpha_3 = k$ through the line $t = 0$.

When $A^2 < 1$, that is, when $|\alpha_3| > 2$, the curve is J-shaped with an infinite ordinate at $t = -A$.

The special case for which $A^2 = 1$ is known in the Pearson system as Type X. When $\alpha_3 = \pm 2$, (22) becomes

$$y = Ke^{\pm t}.$$

This is also known as Laplace's second frequency curve.

Tables of ordinates and areas of the Type III curve have been published by Salvosa in the *Annals of Mathematical Statistics*, vol. 1, no. 2.

A systematic treatment of all the curves in the Pearson system has recently been given in a paper entitled *A New Exposition and Chart for the Pearson System of Frequency Curves* by C. C. Craig, *Annals of Mathematical Statistics*, vol. 7, no. 1, pp. 16–28.

4. Genesis of the Pearson Curves in the Theory of Probability. The differential equation (12) is supposed to have some support in the theory of probability. This claim rests on the assumption that the distribution of statistical material may be likened to *a priori* distributions in certain urn schemata. The method by which (12) is associated with underlying probabilities is started by considering the following problem.

An urn contains n balls of which np are white, so that the probability of drawing a white ball in a single trial is p. The rest of the balls, nq, are black, and the probability of failure to draw a white ball in a single trial is $q = 1 - p$. If s balls are drawn from the urn one at a time with replacements after each draw, what is the probability, $B(x)$, of drawing exactly x white balls and $(s - x)$ black balls?

From the Bernoulli theory it is known that the probabilities of getting $x = 0, 1, 2, \cdots, s$, successes in s trials are given by the successive terms of the binomial

$$(q + p)^s = \sum_{x=0}^{s} B(x) \tag{24}$$

where

$$B(x) = C(s, x)p^x q^{s-x}.$$

Representing the terms $B(x)$ by ordinates y_x, one may plot the $(s + 1)$ points (x, y_x). Through these $(s + 1)$ points one may imagine a curve that can be represented by an analytic function. Since

$$y_x = C(s, x)p^x q^{s-x}$$

and hence

$$y_{x+1} = C(s, x + 1)p^{x+1} q^{s-x-1},$$

we have

$$\frac{y_{x+1}}{y_x} = \frac{sp - px}{qx + q}. \tag{25}$$

From (25) we obtain

$$\frac{y_{x+1} - y_x}{y_{x+1} + y_x} = \frac{sp - q - x}{sp + q + (q - p)x}. \tag{26}$$

Now the mean of any two ordinates (y_x and y_{x+1}) may be considered as approximately equal to the ordinate ($y_{x+1/2}$) midway between them. The slope of the line joining any two points (x, y_x) and $(x + 1, y_{x+1})$ is also approximately equal to the slope of the tangent at the point midway between these two points on the continuous curve. Under these two assumptions, (26) may be written as

$$\frac{D_x y_{x+1/2}}{y_{x+1/2}} = \frac{2(sp - q - x)}{sp + q + (q - p)x}. \tag{27}$$

The right member of this equation is, therefore, the derivative of $\log y$ at the point $(x + \frac{1}{2}, y_{x+1/2})$. At any point (x, y_x) this derivative is

$$\frac{d(\log y)}{dx} = \frac{2\{sp - q - (x - \frac{1}{2})\}}{sp + q + (q - p)(x - \frac{1}{2})}. \tag{28}$$

If $p = q = \frac{1}{2}$, then (28) becomes

$$\frac{d}{dx}(\log y) = \frac{-\left(x - \dfrac{s}{2}\right)}{\left(\dfrac{s+1}{4}\right)},$$

which is of the form

$$\frac{dy}{y} = -\frac{(x - m)\,dx}{a}, \qquad a > 0,$$

and which, upon integration, yields the normal curve

$$y = ke^{-P} \tag{29}$$

where $$P = \frac{(x - m)^2}{2a}.$$

The next step consists in dealing with the case $p \neq q$. From (28) we have

$$-\frac{d}{dx}(\log y) = \frac{\dfrac{q - p}{2} + (x - sp)}{spq + \dfrac{1}{4} + \dfrac{(x - sp)(q - p)}{2}}.$$

If we set

$$\alpha_3 = \frac{q - p}{(spq)^{1/2}} \qquad \text{and} \qquad t = \frac{x - sp}{(spq)^{1/2}}$$

the above equation becomes

$$-\frac{d}{dt}(\log y) = \frac{\dfrac{\alpha_3}{2} + t}{1 + \dfrac{\alpha_3}{2}t + \dfrac{1}{4spq}}. \tag{30}$$

If spq is so large that $1/4spq$ is negligibly small, (30) becomes

$$\frac{d(\log y)}{dt} = -\frac{\dfrac{\alpha_3}{2} + t}{1 + \dfrac{\alpha_3}{2}t} \tag{31}$$

which upon integration yields the Type III curve. It is evident from (31) that this curve approaches the normal curve as a limit as $\alpha_3 \to 0$.

With $p = q$, (28) is of the form (12) when $b = c = 0$. With $p \neq q$, (28) is of the form (12) when $c = 0$. To produce, in the theory of probability, an expression comparable to (12) when both b and c are different from zero it is necessary to consider a more

general urn problem. So far the underlying probability, p, has been constant. If we consider the urn schemata previously described, but remove the restriction of replacements, then the chance of success is not constant from trial to trial but depends upon the results of previous trials. Thus, without replacements, the chances of obtaining exactly $x = 0, 1, 2, \cdots, s$ white balls in a draw of s balls, are given by the successive terms of the hypergeometric series

$$(32) \quad \frac{1}{C(n, s)} \{C(np, 0)C(nq, s) + C(np, 1)C(nq, s - 1) + \cdots + C(np, x)C(nq, s - x) + \cdots + C(np, s)C(nq, 0)\}$$

in which the general term is

$$H(x) = \frac{(np)\,!\,(nq)\,!\,s\,!\,(n - s)\,!}{(np - x)\,!\,(nq - s + x)\,!\,n\,!\,x\,!\,(s - x)\,!}.$$

By representing the terms of this series as ordinates of a frequency polygon, it is possible to show that* the slope, at the mid-point of any side, divided by the ordinate at that point is equal to a fraction whose numerator is a linear function of x and whose denominator is a quadratic function of x. It is clear that (12) gives a general statement of this property.

Since the hypergeometric series is associated with (12) and the Bernoulli series is associated with a special case of (12), *viz.*, when $c = 0$, we should quite naturally expect that the Bernoulli series is a special case of the hypergeometric series. Writing $H(x)$ in the form

$$H(x) = \frac{s\,!}{x\,!\,(s - x)\,!} \times \frac{p\{p - 1/n\} \cdots \{p - (x + 1)/n\}q\{q - 1/n\} \cdots \{q - (s - x - 1)/n\}}{\{1 - 1/n\} \cdots \{1 - (x + 1)/n\}\{1 - x/n\} \cdots \{1 - (s - 1)/n\}}$$

it is obvious that

$$\lim_{n \to \infty} H(x) = C(s, x)p^x q^{s-x} = B(x).$$

When $n = \infty$, there is an infinite supply in the urn, so the probability, p, remains constant from trial to trial without replacements. In other words, sampling from a finite supply with replacements is the same as sampling from an infinite supply without replacements.

* See 1. Elderton, *Frequency Curves and Correlation.*
2. Rietz, *Mathematical Statistics* (Carus Monograph), Chapter III.

5. Further Discussion of the Normal Curve. We will now return to a discussion of the normal curve, giving some proofs which had to be omitted in Part I, and supplying explanations which in one instance or another perhaps had to be read between the lines there.

A. Fitting the Curve. If (29) is to represent an observed distribution, the parameters m, a, and k may be determined by the principle of moments. Equating the kth functional moment to the kth moment of observed data, for $k = 0, 1, 2$, we have three simultaneous equations

$$(33) \qquad \begin{cases} k\displaystyle\int_{-\infty}^{\infty} e^{-(x-m)^2/2a}\, dx = N \\ k\displaystyle\int_{-\infty}^{\infty} e^{-(x-m)^2/2a}\, x\, dx = N\bar{x} \\ k\displaystyle\int_{-\infty}^{\infty} e^{-(x-m)^2/2a}\, x^2\, dx = N\nu_2 \end{cases}$$

in which the parameters are the unknowns.

The solution of these equations can be made to depend upon the integral

$$(34) \qquad \int_{-\infty}^{\infty} e^{-y^2/2a}\, dy = (2\pi a)^{1/2}$$

which is evaluated in Chapter II. Using this result and letting $y = x - m$, the first of equations (33) becomes

$$(a) \qquad k(2\pi a)^{1/2} = N.$$

The second becomes

$$k\int_{-\infty}^{\infty} e^{-y^2/2a}\, y\, dy + km\int_{-\infty}^{\infty} e^{-y^2/2a}\, dy = N\bar{x}.$$

In the above relation, the first integral vanishes because the integrand is an odd function. So, using (34), we have

$$(b) \qquad km(2\pi a)^{1/2} = N\bar{x}.$$

The third integral in (33) may be written in the form

$$k\int_{-\infty}^{\infty} e^{-y^2/2a}y^2\, dy + 2km\int_{-\infty}^{\infty} e^{-y^2/2a}y\, dy + km^2\int_{-\infty}^{\infty} e^{-y^2/2a}\, dy.$$

Upon integrating (by parts) the first integral in the above expression and evaluating the other integrals, we obtain

$$(c) \qquad k\sqrt{2\pi a}(m^2 + a) = N\nu_2.$$

From (a) and (b) we find $m = \bar{x}$. From (a) and (c) we have

$$m^2 + a = \nu_2,$$

and so

$$a = \mu_2 = \sigma^2.$$

Therefore, (29) becomes

$$y = \frac{N}{\sigma\sqrt{2\pi}} e^{-(x-\bar{x})^2/2\sigma^2}. \tag{35}$$

B. Moments. The general moment of odd order of (35) about the mean is given by

$$\mu_{2k+1} = \frac{1}{N}\int_{-\infty}^{\infty} y(x - \bar{x})^{2k+1}\, dx.$$

But the right member vanishes because the integrand is an odd function. Therefore, *all moments of odd order of the normal curve taken about the mean are zero.*

The general moment of even order is

$$\mu_{2k} = \frac{1}{N}\int_{-\infty}^{\infty} y(x - \bar{x})^{2k}\, dx.$$

Integrating the right member by parts, letting $u = (x - \bar{x})^{2k-1}$, the following recursion relation is obtained for even moments

$$\mu_{2k} = (2k - 1)\sigma^2\mu_{2k-2}. \tag{36}$$

Then when $k = 1$, $\mu_2 = \sigma^2$; when $k = 2$, $\mu_4 = 3\mu_2{}^2$; etc.

A recursion formula for the moments in standard units may also be obtained from (19). Under the conditions imposed for Type VII, (19) becomes

$$\alpha_{k+1} = k\alpha_{k-1}, \qquad k = 1, 3, 5, \cdots.$$

Hence,

$$\begin{aligned}
\alpha_2 &= 1\\
\alpha_4 &= 3\\
\alpha_6 &= 1 \cdot 3 \cdot 5\\
&\cdot \quad \cdot \quad \cdot\\
\alpha_{2k} &= 1 \cdot 3 \cdot 5 \cdots (2k - 1)\\
&= \frac{(2k)\,!}{2^k k\,!}.
\end{aligned}$$

C. Quadrature. Some writers use the term *quadrature* for the evaluation of an integral. The definite integral

$$\Phi(t) = \frac{h}{\sqrt{\pi}}\int_0^t e^{-h^2x^2}\,dx, \qquad h = \frac{1}{\sqrt{2}}, \tag{37}$$

is commonly called the *probability integral.* Clearly it is a function of the variable limit t. Although (37) cannot be evaluated in finite form, it can be computed by expanding the integrand into a power series and integrating as many terms as may be needed.

In (37) let $y = hx$. When $x = t$, $y = ht$. So (37) becomes

$$\Phi(t) = \frac{1}{\sqrt{\pi}}\int_0^{ht} e^{-y^2}\,dy. \tag{38}$$

Expanding the integrand of (38) we have

$$e^{-y^2} = 1 - y^2 + \frac{y^4}{2\,!} - \frac{y^6}{3\,!} + \cdots + \frac{y^{2n-2}}{(n-1)\,!} + \cdots.$$

Termwise integration yields the result

$$\frac{1}{\sqrt{\pi}}\int_0^y e^{-y^2}\,dy = \frac{1}{\sqrt{\pi}}\left\{y - \frac{y^3}{3} + \frac{y^5}{10} - \frac{y^7}{42} + \frac{y^9}{216} - R\right\}, R < \frac{y^{11}}{1320}. \tag{39}$$

This series converges for all values of y, and the error made in stopping at any term is numerically less than the first term neglected. For small values of y it converges rapidly and is a satisfactory method for computing when $y \leq 1$.

But for large values of y, (39) converges too slowly to be practical; too many terms are required. It is therefore important to obtain an expansion in descending powers of y. To this end write

$$\int_0^y e^{-y^2}\,dy = \int_0^\infty e^{-y^2}\,dy - \int_y^\infty e^{-y^2}\,dy$$

$$= \frac{\sqrt{\pi}}{2} - \int_y^\infty e^{-y^2}\,dy, \tag{40}$$

and

$$\int_y^\infty e^{-y^2}\,dy = \int_y^\infty \frac{1}{y}ye^{-y^2}\,dy.$$

Integrating the last integral by parts we obtain

$$-\frac{e^{-y^2}}{2y}\Big]_y^\infty - \frac{1}{2}\int_y^\infty \frac{e^{-y^2}}{y^2}\,dy\,.$$

Integrating successively by parts gives the result

$$\int_y^\infty e^{-y^2}\,dy = \frac{e^{-y^2}}{2y}\left\{1 - \frac{1}{2y^2} + \frac{3}{4y^4} - \frac{3\cdot 5}{8y^6} + \cdots\right\}. \tag{41}$$

From (40) and (41) we have the final result

$$\frac{1}{\sqrt{\pi}}\int_0^y e^{-y^2}\,dy = .5 - \frac{e^{-y^2}}{2y\sqrt{\pi}}\left\{1 - \frac{1}{2y^2} + \frac{3}{4y^4} - \frac{3\cdot 5}{8y^6} + \cdots + (-1)^n T_{n+1} + \cdots\right\} \tag{42}$$

where

$$T_{n+1} = \frac{1\cdot 3\cdot 5\cdots(2n-1)}{2^n y^{2n}}$$

and $(n + 1)$ is the number of the term. The series in (42) is called an *asymptotic* or semi-convergent series; it converges until a minimum term is reached and then diverges. The general term T_{n+1} decreases so long as $n \leq y^2$. But after the integrations by parts have been performed so many times that $n > y^2$, T_{n+1} increases. Of course the integrations should not be carried further. The value obtained by using the series in (42) will differ from the true value by less than the last term retained.

Tables of (37) may be computed by means of (39) for $y(= ht) \leq 1$ and by (42) for $y > 1$. Such tables were computed long ago and are available in many places.

Example. Evaluate (37) for $t = 3$ and check the result with the value of $\int_0^t \phi(t)\,dt$ for $t = 3$ given in the tables in the Appendix.

Solution. Since $y = t/\sqrt{2}$ we are to evaluate (42) for $y = 3/\sqrt{2}$. Substituting this value in (42) we have

$$\begin{aligned} .5 - \frac{e^{-9/2}}{2\sqrt{\pi}}\frac{\sqrt{2}}{3}&\left\{1 - \frac{1}{9} + \frac{3}{81} - \frac{15}{729} + \frac{105}{6561}\right\} \\ &= .5 - \frac{1}{3}\phi(3)\cdot(.9213) \\ &= .5 - .00136 = .49864. \end{aligned}$$

The value given in the tables is .49865.

6. The Gram-Charlier Series. If a function $f(x)$ gives only a rough approximation to a frequency distribution, a more accurate representation may be obtained by using the first few terms of the series

$$F(x) = c_0 f(x) + c_1 f^{(1)}(x) + c_2 f^{(2)}(x) + \cdots + c_n f^{(n)}(x) + \cdots \tag{43}$$

where $f(x)$, called the "generating function," gives a first approximation to the given distribution, and $f^{(n)}(x)$ is the nth derivative of $f(x)$ with respect to x.

It should be observed that series representation is also involved in the Pearson system. For, suppose the differential equation underlying that system is written in the form

$$\frac{dy}{dx} = \frac{y(a - x)}{f(x)}.$$

Then if it be assumed that $f(x)$ is expressible as a power series which is so rapidly convergent that the first few terms are sufficient, we have the form given in (12). In the Pearson system the series occurs in the differential equation of the function whereas in the Gram-Charlier system it occurs in the function itself.

If in (43) the normal curve is taken as the generating function then $F(x)$ is known as the Gram-Charlier Type A series. In discussing this series no essential loss of generality is suffered by using standard units. Thus we may write

$$F(t) = c_0\phi(t) + c_1\phi^{(1)}(t) + c_2\phi^{(2)}(t) + \cdots + c_n\phi^{(n)}(t) + \cdots \tag{44}$$

where $\phi(t)$ is defined in (21). The moments of $F(t)$ are defined by

$$\alpha_n = \int_{-\infty}^{\infty} F(t)t^n\,dt \tag{45}$$

and it follows that $\alpha_0 = 1,\ \alpha_1 = 0,\ \alpha_2 = 1.$

The coefficients c_n in (44) may be expressed in terms of the moments α_n, because the functions $\phi^{(n)}(t)$ and the Hermite polynomials $H_m(t)$ defined by the relation

$$\phi^{(m)}(t) = (-1)^m H_m(t)\phi(t) \tag{46}$$

form a biorthogonal system. That is

$$\int_{-\infty}^{\infty} \phi^{(n)}(t)H_m(t)\,dt = 0 \qquad \text{for } m \neq n, \tag{47}$$

$$\int_{-\infty}^{\infty} \phi^{(n)}(t)H_m(t)\,dt = (-1)^n n\,! \qquad \text{for } m = n. \tag{48}$$

Proofs of (47) and (48) are available in the literature* and will be omitted here. The recursion relation

$$H_{n+1}(t) = tH_n(t) - nH_{n-1}(t) \tag{49}$$

can be established.† By differentiating $\phi(t)$ we find from (46) that $H_1 = t$ and since $H_0 = 1$ we can use (49) for $n \geq 1$.

To make use of the biorthogonal property noted above we multiply both members of (44) by $H_n(t)$ and integrating, under the assumption that the series is uniformly convergent, we obtain

$$\int_{-\infty}^{\infty} F(t)H_n(t)\,dt = c_n \int_{-\infty}^{\infty} \phi^{(n)}(t)H_n(t)\,dt = c_n(-1)^n n\,! \tag{50}$$

since all terms of the right member vanish except the one with the coefficient c_n. Hence from (50) we have

$$c_n = \frac{(-1)^n}{n\,!}\int_{-\infty}^{\infty} F(t)H_n(t)\,dt. \tag{51}$$

From (51), (49), and (45) we obtain the following results:

$$c_0 = \int_{-\infty}^{\infty} F(t)\,dt = 1$$

$$c_1 = \int_{-\infty}^{\infty} F(t)t\,dt = 0$$

$$c_2 = \frac{1}{2}\int_{-\infty}^{\infty} F(t)(t^2 - 1)\,dt = 0$$

$$c_3 = \frac{1}{3\,!}\int_{-\infty}^{\infty} F(t)(-t^3 + 3t)\,dt = -\frac{\alpha_3}{3\,!}$$

$$c_4 = \frac{1}{4\,!}\int_{-\infty}^{\infty} F(t)(t^4 - 6t^2 + 3)\,dt = \frac{\alpha_4 - 3}{4\,!}.$$

We have, therefore,

$$F(t) = \phi(t) - \frac{\alpha_3}{6}\phi^{(3)}(t) + \frac{\alpha_4 - 3}{4\,!}\phi^{(4)}(t) + \cdots \tag{52}$$

and

$$F(x) = \frac{N}{\sigma}F(t).$$

The values of $\phi(t)$, of its integral, and of its second to eighth derivative, are given to five places of decimals in *Glover's Tables.*

* See Rietz, *Mathematical Statistics*, pp. 165–168.

† See Levy and Roth, *Elements of Probability*. Oxford. 1936.

Exercises

1. Prove that the points of inflection of the normal curve are equidistant from the mode. What are the coördinates of these points?

2. If x has the distribution function $y = f(x)$, with total frequency 1, the mean deviation, M, about the value v is defined by

$$M = \int_{-\infty}^{\infty} |x - v| f(x)\, dx.$$

Prove that M is a minimum when v is the median, that is, when the ordinate at $x = v$ bisects the area under $y = f(x)$.

Solution. We may write the expression for M in the form

$$M = \int_{-\infty}^{v} (v - x) f(x)\, dx + \int_{v}^{\infty} (x - v) f(x)\, dx.$$

It is shown in treatises on advanced calculus that if

$$H(\theta) = \int_{a}^{b} f(x, \theta)\, dx,$$

θ being a parameter and a and b being functions of θ, then

$$\frac{dH}{d\theta} = \int_{a}^{b} \frac{\partial f}{\partial \theta}\, dx - f(a, \theta)\frac{da}{d\theta} + f(b, \theta)\frac{db}{d\theta}.$$

Therefore, differentiating M with respect to v and equating the result to zero, we have

$$\int_{-\infty}^{v} f(x)\, dx - \int_{v}^{\infty} f(x)\, dx = 0.$$

So M is a minimum when $\int_{-\infty}^{v} f(x)\, dx = \int_{v}^{\infty} f(x)\, dx$, that is, when the partial areas to the right and left of v are equal. (It is left to the student to show that M is actually a minimum when $dM/dv = 0$.)

3. Prove that the relation between the mean deviation (about the mean) and the standard deviation of the normal curve (in arbitrary units) is

$$M = (2/\pi)^{1/2}\sigma = .798\sigma, \text{ approximately.}$$

Hint. By definition,

$$M = \frac{1}{N}\int_{-\infty}^{\infty} y\, |x - \bar{x}|\, dx = \sigma \int_{-\infty}^{\infty} \phi(t)\, |t|\, dt = 2\sigma \int_{0}^{\infty} \phi(t) t\, dt.$$

4. Suppose x is distributed in accord with the frequency curve $y = Ce^{-x/a}$, $0 \leq x \leq \infty$, a being a positive constant and C being determined by the condition that the area under the curve is N. Evaluate ν_k successively for $k = 1, 2, 3, 4$. Then find μ_k for $k = 2, 3, 4$, and finally obtain the values $\bar{x} = a$, $\sigma = a$, $\alpha_3 = 2$, $\alpha_4 = 9$.

5. Given $f(x) = Cx^{n-1}e^{-x}$, $0 \leq x \leq \infty$, where C is determined by the condition that the area under the curve is unity. Evaluate ν_k for $k = 1$ to 4, μ_k for $k = 2$ to 4, and α_k for $k = 3, 4$. Show that α_3 and α_4 satisfy the criterion $2\alpha_4 - 3\alpha_3^2 - 6 = 0$.

6. State the differential equation underlying the Pearson system of frequency curves and derive the equation of the normal curve as a special solution of this equation. Evaluate the constant of integration so that the area under the curve is unity.
7. Discuss the Type III curve.
8. Show that y in (22) vanishes when $t = -A$ and $t = \infty$.
9. Read Chapter III of the *Carus Monograph on Mathematical Statistics* by H. L. Rietz.
10. Explain how the probability integral (37) may be evaluated for, (*a*) small values of t, (*b*) large values of t.
11. Evaluate (37) for (*a*) $t = \sqrt{2}/2$, (*b*) $t = 2\sqrt{2}$.
12. Consult the reference cited for the proofs of (47) and (48) and give a report on them.
13. By successive differentiation of $\phi(t)$ evaluate $H_m(t)$ from (46) for $m = 1, 2, 3, 4$. Check your results with (49) for $n = 1, 2, 3$.
14. Making use of the biorthogonal property of Hermite polynomials and derivatives of the normal curve, derive the values of c_n, $n = 0$ to 4, in the Type A series.
15. Taking $t = 0, \pm 1, \pm 2, \pm 3$, plot (52) on the same axes when (*a*) $\alpha_3 = 0$ and $\alpha_4 = 3$, (*b*) $\alpha_3 = -1.2$ and $\alpha_4 = 3$, (*c*) $\alpha_3 = -1.2$ and $\alpha_4 = 4.2$. In (*b*) if $\alpha_3 = 1.2$, what effect would this have on the curve?

CHAPTER IV

JOINT DISTRIBUTIONS OF TWO VARIABLES. THE NORMAL CORRELATION SURFACE

1. Fundamental Notions. Definitions of a frequency function of one variable and the associated notion of probability were given in Chapter III. Corresponding definitions will now be given for an arbitrary probability distribution of two variables. The continuous variables (x, y) have the joint probability function $f(x, y)$ if the double integral of $f(x, y)$ over a region of the (x, y)-plane measures the relative frequency of occurrence of pairs of values (x, y) in that region. It will be understood that $f(x, y)$ is continuous, single-valued, and non-negative. If values of (x, y) are restricted to a finite region we define $f(x, y)$ to be identically zero outside that region. In the extended region of definition, we have

$$(1) \qquad \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)\, dy\, dx = 1.$$

Geometrically, this means that the volume under the surface represented by $z = f(x, y)$ is unity. Then $f(x, y)\, dy\, dx$ is the probability that simultaneously x lies in the interval $(x, x + dx)$ and y lies in the interval $(y, y + dy)$. Consequently,

$$(2) \qquad \int_{a}^{b}\int_{c}^{d} f(x, y)\, dy\, dx$$

represents the probability that x lies between a and b at the same time that y lies between c and d.

We shall distinguish between two cases: (a) when the variables are independent in the probability sense, and (b) when they are correlated. Let the probability be $g(x)\, dx$ that x occurs in dx for all y's. Then integrating over all admissible values of y, we have

$$(3) \qquad g(x)\, dx = dx \int_{-\infty}^{\infty} f(x, y)\, dy.$$

It is clear that the integral in (3) gives $g(x)$ because the relative frequency of occurrence of x in any interval (a, b) is the relative

frequency of pairs (x, y) belonging to the strip of the xy-plane for which $a < x < b$, and this is

$$\int_a^b \int_{-\infty}^{\infty} f(x, y)\, dy\, dx = \int_a^b g(x)\, dx.$$

Similarly, if $h(y)\, dy$ is the probability that y occurs in dy for all assignments of x, we have

$$h(y)\, dy = dy \int_{-\infty}^{\infty} f(x, y)\, dx. \tag{4}$$

In accordance with convention we shall call $g(x)$ and $h(y)$ the *marginal* distributions.

The independence of x and y is characterized by the following

DEFINITION. *The variables x and y are independent when $f(x, y) \equiv g(x)h(y)$. If $f(x, y)$ cannot be expressed identically as the product of the marginal distributions, then x and y are said to be correlated.*

2. Moments. Let the general product moment about the common origin of x and y be defined as follows:

$$\nu_{mn} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) x^m y^n\, dy\, dx. \tag{5}$$

If $m = 0$ and $n = 1$, we have

$$\nu_{01} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) y\, dy\, dx. \tag{6}$$

Let $f(x, y)$ be a function in which the order of integration may be interchanged. Then ν_{01} becomes

$$\int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} f(x, y)\, dx \right] y\, dy = \int_{-\infty}^{\infty} h(y) y\, dy,$$

which is the mean, $\bar{y}$, of the y's. Similarly, the mean of the x's is

$$\nu_{10} = \bar{x} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) x\, dy\, dx = \int_{-\infty}^{\infty} g(x) x\, dx. \tag{7}$$

We will now define the general product moment about the means $(\bar{x}, \bar{y})$ as follows:

$$\mu_{mn} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \bar{x})^m (y - \bar{y})^n f(x, y)\, dy\, dx. \tag{8}$$

When $m = n = 1$, we have

$$\mu_{11} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x - \bar{x})(y - \bar{y})f(x, y)\, dy\, dx, \tag{9}$$

which is styled the *co-variance* of the joint distribution.

When $m = 2$ and $n = 0$, we have the *variance of* x,

$$\begin{aligned} \mu_{20} &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x - \bar{x})^2 f(x, y)\, dy\, dx \\ &= \int_{-\infty}^{\infty} (x - \bar{x})^2 g(x)\, dx \\ &= \sigma_x^2. \end{aligned} \tag{10}$$

Similarly, when $m = 0$ and $n = 2$, we have the *variance of* y,

$$\begin{aligned} \mu_{02} &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (y - \bar{y})^2 f(x, y)\, dy\, dx \\ &= \int_{-\infty}^{\infty} (y - \bar{y})^2 h(y)\, dy \\ &= \sigma_y^2. \end{aligned} \tag{11}$$

It is left as an exercise for the student to show that

$$\begin{cases} \mu_{11} = \nu_{11} - \nu_{10}\nu_{01}, \\ \mu_{20} = \nu_{20} - \nu_{10}^2. \end{cases} \tag{12}$$

The coefficient of correlation between x and y, denoted by ρ_{xy}, is defined by

$$\rho_{xy} = \frac{\mu_{11}}{\sigma_x \sigma_y}. \tag{13}$$

3. Regression. If y has been assigned in the joint probability function $f(x, y)$, the probability that x will lie in an infinitesimal interval is

$$\frac{f(x, y)}{h(y)}\, dx.$$

Thus, when y is fixed,

$$\int_{-\infty}^{\infty} \frac{f(x, y)}{h(y)}\, dx = 1,$$

and so $f(x, y)/h(y)$ is the probability function of x for a fixed y. It may be called the probability density representing a y array of x's.

Likewise, if we fix x the probability density for an x array of y's is given by $f(x, y)/g(x)$, since

$$\int_{-\infty}^{\infty} \frac{f(x, y)}{g(x)} dy = 1$$

when x is fixed.

The notion of arrays may be made more concrete by thinking of a joint distribution of the heights and weights of men. If x refers to weight and y to height, then an example of an x array of y's is the distribution of the heights of all men who weigh 150 pounds, and the weights of all men who are six feet tall is an example of a y array of x's.

The mean of an x array of y's is

$$\bar{y}_x = \int \frac{yf(x, y)}{g(x)} dy \tag{14}$$

where the integration is performed over all values in the array defined by x. Similarly, the mean of a y array of x's is

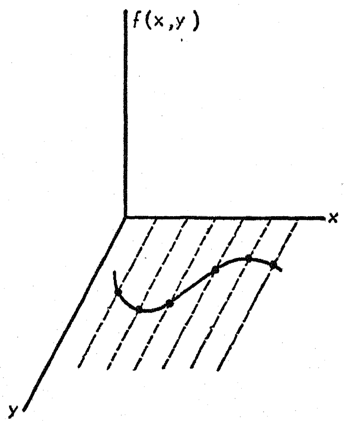
$$\bar{x}_y = \int \frac{xf(x, y)}{h(y)} dx \tag{15}$$

integrated over all x's in an array for a fixed y.

The *variance in an x array of y*'s is given by

$$\int (y - \bar{y}_x)^2 \frac{f(x, y)}{g(x)} dy \tag{16}$$

integrated over all values in the array fixed by x. Similarly, the *variance in a y array of x*'s is



$$\int (x - \bar{x}_y)^2 \frac{f(x, y)}{h(y)} dx \tag{17}$$

integrated over all values in the array fixed by y.

Taking different x arrays of y's fixes the mean points $\bar{y}_x$ and as x varies continuously we get the locus of these means which is called the *regression curve of y on x*. Its equation is given by (14) where now, of course, x is a variable. Similarly, (15) gives the *regression curve of x on y*. Of particular interest and use are the cases in which these

regression curves are straight lines. If the equation of the regression curve of y on x is of the form

$$\bar{y}_x = Ax + B,$$

then the regression of y on x is said to be linear. Similarly, if the equation of the regression curve of x on y is of the form

$$\bar{x}_y = Cy + D,$$

then the regression of x on y is said to be linear. If one regression system is linear the other is not necessarily linear.

Let us now consider the implications of linear regression on the joint probability function $f(x, y)$ and the marginal totals $g(x)$ and $h(y)$. Consider

$$\bar{y}_x = \int_{-\infty}^{\infty} \frac{yf(x, y)}{g(x)}\, dy = Ax + B.$$

or

$$\int_{-\infty}^{\infty} yf(x, y)\, dy = Axg(x) + Bg(x). \tag{18}$$

Integrating each side of (18) with respect to x, and remembering that we may interchange the order of integration, we have

$$\int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} yf(x, y)\, dy\right] dx = A\int_{-\infty}^{\infty} xg(x)\, dx + B\int_{-\infty}^{\infty} g(x)\, dx,$$

or

$$\nu_{01} = A\nu_{10} + B. \tag{19}$$

Multiplying each side of (18) by x and integrating with respect to x, we have

$$\int_{-\infty}^{\infty} x\left[\int_{-\infty}^{\infty} yf(x, y)\, dy\right] dx = A\int_{-\infty}^{\infty} x^2 g(x)\, dx + B\int_{-\infty}^{\infty} xg(x)\, dx.$$

Since the left member is

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf(x, y)\, dy\, dx = \nu_{11},$$

we have

$$\nu_{11} = A\nu_{20} + B\nu_{10}. \tag{20}$$

A simultaneous solution of (19) and (20) yields

$$A = \frac{\nu_{11} - \nu_{10}\nu_{01}}{\nu_{20} - \nu_{10}^2} = \frac{\mu_{11}}{\mu_{20}} = \rho\frac{\sigma_y}{\sigma_x},$$

$$B = \nu_{01} - \nu_{10}\frac{\sigma_y}{\sigma_x}\rho$$

$$= \bar{y} - \bar{x}\frac{\sigma_y}{\sigma_x}\rho.$$

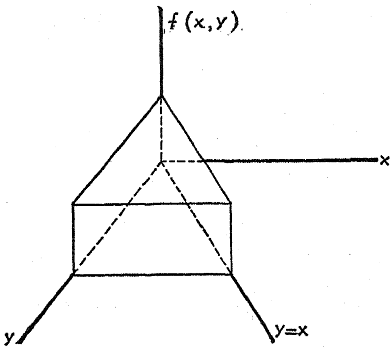
Therefore the equation of the *line of regression of* y *on* x becomes

$$\bar{y}_x - \bar{y} = \rho\frac{\sigma_y}{\sigma_x}(x - \bar{x}). \tag{21}$$

In an analogous manner, if the *regression of* x *on* y is linear the *regression line* has the equation

$$\bar{x}_y - \bar{x} = \rho\frac{\sigma_x}{\sigma_y}(y - \bar{y}). \tag{22}$$

The quantities $A = \rho(\sigma_y/\sigma_x)$ and $C = \rho(\sigma_x/\sigma_y)$ are called the *regression coefficients.* It is obvious that their product is ρ^2.



Example. Given

$$f(x, y) = \frac{2}{a^2}, \qquad \begin{array}{l} 0 \leq x \leq y, \\ 0 \leq y \leq a, \end{array}$$

as the joint probability function of two variables x and y. Find (i) the marginal totals $g(x)$ and $h(y)$; (ii) the mean and variance of each of the marginal totals, *i.e.*, ν_{10} and $\sigma_x^2 = \mu_{20}$ for $g(x)$, ν_{01} and $\sigma_y^2 = \mu_{02}$ for $h(y)$; (iii) the equations of the regression curves of y on x and of x on y, $\bar{y}_x$ and $\bar{x}_y$; (iv) the correlation coefficient ρ.

Solutions. The volume under the surface represented by the given function is unity. Thus

$$\int_0^a \int_0^y \frac{2}{a^2}\,dx\,dy = \frac{2}{a^2}\int_0^a y\,dy = 1.$$

The surface is shown above.

(i) The marginal totals are

$$g(x) = \int_x^a \frac{2}{a^2}\,dy = \frac{2}{a^2}(a - x)$$

$$h(y) = \int_0^y \frac{2}{a^2}\,dx = \frac{2y}{a^2}.$$

(ii) The means are

$$\nu_{10} = \bar{x} = \int_0^a x \frac{2}{a^2}(a - x)\, dx = \frac{a}{3}$$

$$\nu_{01} = \bar{y} = \int_0^a y \frac{2y}{a^2}\, dy = \frac{2a}{3}.$$

Since

$$\nu_{20} = \int_0^a x^2 \frac{2}{a^2}(a - x)\, dx = \frac{a^2}{6}$$

$$\nu_{02} = \int_0^a y^2 \frac{2}{a^2} y\, dy = \frac{a^2}{2}$$

the variances are

$$\mu_{20} = \sigma_x^2 = \frac{a^2}{6} - \frac{a^2}{9} = \frac{a^2}{18}$$

$$\mu_{02} = \sigma_y^2 = \frac{a^2}{2} - \frac{4a^2}{9} = \frac{a^2}{18}.$$

(iii) The regression lines are

$$\bar{y}_x = \int_x^a y \frac{2/a^2}{2(a - x)/a^2}\, dy = \frac{a + x}{2}$$

$$\bar{x}_y = \int_0^y x \frac{2/a^2}{2y/a^2}\, dx = \frac{y}{2}.$$

(iv) From the equations of the regression lines it follows that $\rho^2 = \frac{1}{4}$ and $\rho = \frac{1}{2}$ since $\rho(\sigma_y/\sigma_x)$ is positive.

4. The Standard Error of Estimate. We have seen that the probability density in an x array of y's is $f(x, y)/g(x)$. Then the variance $s_{y \cdot x}^2$ within such an array is

$$s_{y.x}^2 = \int_{-\infty}^{\infty} (y - \bar{y}_x)^2 \frac{f(x, y)}{g(x)}\, dy.$$

The mean, over all x arrays, of values of $s_{y.x}^2$ weighted with the marginal distribution of x is denoted by S_y^2, and S_y is called the *standard error of estimate*. We will now show that $S_y^2 = \sigma_y^2(1 - \rho^2)$. By definition,

$$S_y^2 = \int_{-\infty}^{\infty} g(x) s_{y\,x}^2\, dx$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (y - \bar{y}_x)^2 f(x, y)\, dy\, dx.$$

Using the value of $\bar{y}_x$ given in (21) the above expression becomes

$$S_y^2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left\{ y - \bar{y} - \rho\frac{\sigma_y}{\sigma_x}(x - \bar{x})\right\}^2 f(x, y)\, dy\, dx$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left\{ (y - \bar{y})^2 - 2\rho\frac{\sigma_y}{\sigma_x}(y - \bar{y})(x - \bar{x}) + \rho^2\frac{\sigma_y^2}{\sigma_x^2}(x - \bar{x})^2\right\} f(x, y)\, dy\, dx,$$

and the right member simplifies so that we have the result

$$S_y^2 = \sigma_y^2(1 - \rho^2).$$

From this result it follows that

$$-1 \leq \rho \leq 1.$$

5. The Normal Correlation Surface. We shall now consider a joint probability function of special interest. The normal correlation surface is defined by the following function

$$f(x, y) = Ke^{-P}, \tag{23}$$

where

$$P = \frac{1}{2(1 - \rho^2)}\left\{\frac{x^2}{\sigma_x^2} - \frac{2\rho xy}{\sigma_x\sigma_y} + \frac{y^2}{\sigma_y^2}\right\},$$

$$\frac{1}{K} = 2\pi\sigma_x\sigma_y(1 - \rho^2)^{1/2},$$

$$-\infty \leq x \leq \infty, \qquad -\infty \leq y \leq \infty,$$

and the variables x and y have the origin of their reference system at their respective means, that is,

$$\left\{\begin{aligned} \bar{x} &= \int_{-\infty}^{\infty} xg(x)\, dx = 0, \\ \bar{y} &= \int_{-\infty}^{\infty} yh(y)\, dy = 0. \end{aligned}\right. \tag{24}$$

These conditions (24) may be imposed without essential loss of generality and will simplify the algebraic discussion.

The marginal distribution of x is given by

$$
\begin{aligned}
g(x) &= \int_{-\infty}^{\infty} f(x, y)\, dy \\
&= Ke^{-x^2/2\sigma_x^2}\int_{-\infty}^{\infty} e^{-[y/\sigma_y-\rho x/\sigma_x]^2/2(1-\rho^2)}\, dy \\
&= Ke^{-x^2/2\sigma_x^2}\int_{-\infty}^{\infty} e^{-y'^2/2(1-\rho^2)}\sigma_y\, dy' \\
&= Ke^{-x^2/2\sigma_x^2}\{2\pi(1-\rho^2)\}^{1/2}\sigma_y. \\
&= \frac{1}{\sigma_x\sqrt{2\pi}} e^{-x^2/2\sigma_x^2}. \qquad (25)
\end{aligned}
$$

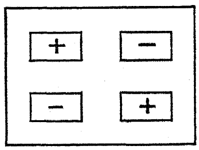
Similarly, the marginal distribution of y is

$$
\begin{aligned}
h(y) &= \int_{-\infty}^{\infty} f(x, y)\, dx \\
&= \frac{1}{\sigma_y\sqrt{2\pi}} e^{-y^2/2\sigma_y^2}.
\end{aligned} \qquad (26)
$$

Hence we may state

Theorem I. *If two variables are normally correlated, each variable is normally distributed in its marginal totals.*

That the converse is not necessarily true is shown by the following illustration. Consider a clay model of a normal correlation surface such that its marginal totals are necessarily normal distributions by the above theorem. Quantities of the clay can be redistributed by piling up in certain spots the clay that is scooped out in other spots in such a way that the marginal totals are not disturbed. It is obvious that the resulting surface is not one that is defined by (23).



Other interesting properties of normally correlated variables are described by the following theorems.

Theorem II. *The regression systems of a normal correlation surface are linear.*

The proof is a matter of integration. Let us find the probability function of an x array of y's. By definition, this is given by $f(x, y)/g(x)$. To get the mean of such an array we must multiply

its probability distribution by y and integrate over all values of y in the array. Thus we have

$$\begin{aligned}\bar{y}_x &= \int_{-\infty}^{\infty} \frac{yf(x, y)\,dy}{g(x)} \\ &= \int_{-\infty}^{\infty} \frac{1}{\sigma_y\{2\pi(1-\rho^2)\}^{1/2}} e^{-[y/\sigma_y-\rho(x/\sigma_x)]^2/2(1-\rho^2)} y\,dy \\ &= \frac{x\rho\sigma_y}{\sigma_x}.\end{aligned}$$

In the exercises at the end of the chapter the student is asked to verify the above result. If x is allowed to vary over the arrays, it is evident that the locus of the means of the x arrays of y's is the line

$$\bar{y}_x = \frac{x\rho\sigma_y}{\sigma_x}. \tag{27}$$

In a similar way the mean of a y array of x's is given by

$$\begin{aligned}\bar{x}_y &= \int_{-\infty}^{\infty} \frac{xf(x, y)\,dx}{h(y)} \\ &= \frac{y\sigma_x\rho}{\sigma_y},\end{aligned}$$

and this lies on the regression line

$$\bar{x}_y = \frac{y\rho\sigma_x}{\sigma_y}. \tag{28}$$

While it is an intrinsic property of a normal correlation surface that both regressions are linear, one should not infer that this is characteristic of joint probability functions in general. One or both or neither of the regression systems of an arbitrary distribution function may be linear. The student will observe that the definition of the correlation coefficient did not involve the condition that $f(x, y)$ was normal nor that regression was linear. Although the *definition* of a correlation coefficient does not require linear regression, nevertheless the correlation coefficient may fail to measure the correlation in the case of appreciable non-linear regression.

Theorem III. *If x and y are normally correlated, then each array is a normal distribution with constant variance S_y^2 from one array of y's to another and constant variance S_x^2 from one array of x's to another.*

The proof consists in exhibiting the distribution function for an x array of y's and for a y array of x's. Thus, for the first case we have

$$\frac{f(x, y)}{g(x)} = \frac{1}{\sqrt{2\pi}S_y} e^{-[y-\rho x(\sigma_y/\sigma_x)]^2/2S_y^2} \tag{29}$$

where $S_y^2 = \sigma_y^2(1 - \rho^2)$. Evidently, this is a normal distribution with variance S_y^2 which is independent of x and therefore is constant over all x arrays. It is left as an exercise for the student to give the companion proof for the arrays in the y direction.

When the variance is constant over the arrays in the x direction the regression system of y on x is said to be *homoscedastic* (equally scattered). Similarly for the y direction. A geometrical representation of a normal correlation surface is given in Part I, § 18 of Chapter VIII.

6. Limiting Forms. Suppose a plane is passed through the surface defined by (23) parallel to the xy-plane. Analytically, this means that we let $f(x, y) = c$ where c is some constant less than the maximum value of the function, that is, we take $0 < c < K$ to insure a real intersection. We obtain

$$\frac{x^2}{\sigma_x^2} - \frac{2\rho xy}{\sigma_x\sigma_y} + \frac{y^2}{\sigma_y^2} = \lambda^2, \tag{30}$$

where

$$\lambda^2 = 2(1 - \rho^2) \log_e \frac{K}{c} \tag{30a}$$

which is obviously not negative. Thus the points (x, y) for which the probability density is constant lie on an ellipse.

It is easier to study (30) if we transform the variables to standard units by letting $t_x = x/\sigma_x$ and $t_y = y/\sigma_y$. Then (30) becomes

$$t_x^2 - 2\rho t_x t_y + t_y^2 = \lambda^2. \tag{31}$$

The cross-product term will vanish under the transformations

$$t_x = u \cos \theta - v \sin \theta$$

$$t_y = u \sin \theta + v \cos \theta$$

when $\theta = \pi/4$. So the required rotation formulas are

$$t_x = \frac{u - v}{(2)^{1/2}} \quad \text{and} \quad t_y = \frac{u + v}{(2)^{1/2}}. \tag{32}$$

Applying these to (31) we obtain

$$u^2(1 - \rho) + v^2(1 + \rho) = \lambda^2 \tag{33}$$

which may be written in the standard form

$$\frac{u^2}{a^2} + \frac{v^2}{b^2} = 1 \tag{34}$$

where

$$a^2 = \frac{\lambda^2}{1 - \rho} \qquad \text{and} \qquad b^2 = \frac{\lambda^2}{1 + \rho}.$$

The eccentricity of the ellipse (34) is $(1 - b^2/a^2)^{1/2} = [2\rho/(1 + \rho)]^{1/2}$. We see that $b \to a$ as $\rho \to 0$. When $\rho = 0$, $b = a = \lambda$. Then (34) would be a circle, and (23) would be a surface of revolution if the variables were expressed in standard units. When $\rho = 1$, it follows from (33) and (30a) that $v = 0$. From (32) it is seen that the line $v = 0$ is the same as $t_y = t_x$, and the ellipse has degenerated into a straight line. The surface then shrinks into a normal curve in the plane $t_y = t_x$.

7. Tetrachoric Correlation. The word *tetrachoric* refers to a 2×2 fold table. Suppose N objects are classified according as they possess one or both or neither of two qualitative traits or attributes which may, for convenience, be denoted by I and II. Such a classification will yield a four fold table as shown in Table 4,

TABLE 4

	Not II	*II*	*Total*
Not I	a	b	$a + b$
I	c	d	$c + d$
Total	$a + c$	$b + d$	N

where $a + b + c + d = N$, the four classes being mutually exclusive but not necessarily exhaustive. The attributes may sometimes admit also of quantitative measurement but we are considering only the case where they are classified in dichotomy, such as " tall " and " not tall," " male " and " female," " alive " and " dead," " good " and " bad," " dull " and " not dull," etc. An example is the follow-

ing classification of 26,287 children where attribute I is dullness and attribute II is developmental defects.

TABLE 5. (K. Pearson, *Tables*, p. li)

	Without Defects	*With Defects*	*Totals*
Not Dull	22,793	1,140	24,213
Dull	1,186	888	2,074
Totals	23,979	2,308	26,287

The problem in such classifications is to measure the intensity of association between the two attributes in the set. Let us suppose that our data had been given initially so that a fine division into many cells was possible and that the result would have presented a normal correlation surface. If this surface were then divided into four cells by planes $x = h$ and $y = k$ to yield the relative frequencies observed, then the correlation coefficient that characterizes this normal correlation surface is called *tetrachoric r*. It will be denoted by r_t.

K. Pearson has given a method and tables for determining r_t. (Cf. *Tables for Statisticians and Biometricians*, Part I.) The procedure may be indicated by the following diagram and skeleton solution for our example, Table 5. (The details will be found in the reference cited.)

Solution of Example. (See Figure 11, page 76.)

$$\int_{-\infty}^{h} = \frac{2074}{26287} = .078{,}898, \qquad h = 1.413;$$

$$\int_{-\infty}^{k} = \frac{2308}{26287} = .087{,}800, \qquad k = 1.354.$$

Entering Pearson's *Tables* for the above values of h and k and interpolating, it is found that $r_t = .652$.

The determination of r_t by Pearson's method is rather tedious when $.2 \leq |r_t| \leq .8$. This burden has been lifted by two fairly recent publications. Camp has given in his text (pp. 307–310) an ingenious and simple method for approximating r_t. His scheme is interesting from the mathematical as well as the practical

point of view. In *Computing Diagrams for the Tetrachoric Correlation Coefficient* by Thurstone *et al.* (available at the University of Chicago Bookstore), a useful approximation to r_t can be determined by inspection.

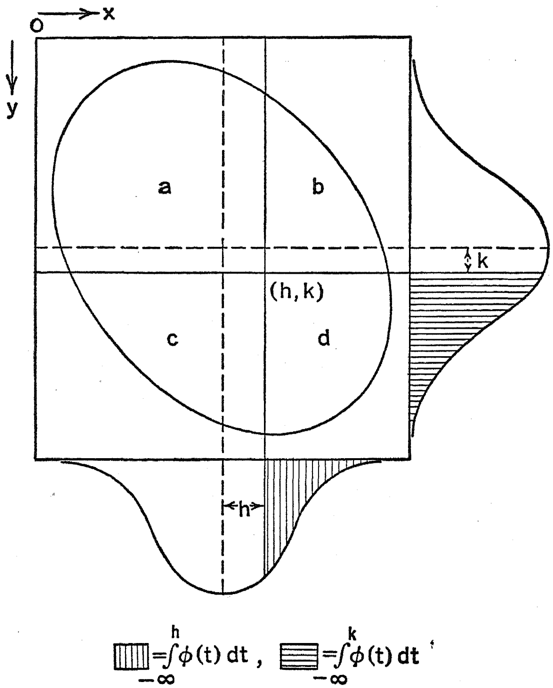


FIG. 11

Exercises

1. Show that the definition of ρ may be written in the form

$$\rho = \frac{1}{\sigma_x\sigma_y}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf(x, y)\, dy\, dx - \bar{x}\bar{y}.$$

2. Given that $f(x, y) = 2/a^2$, $0 \leq x \leq y$, $0 \leq y \leq a$. Show that both regression systems are linear. Evaluate ρ.

3. Derive (22).

4. Prove that the area of the ellipse (30) is $\pi\lambda^2\sigma_x\sigma_y/(1 - \rho^2)^{1/2}$.

5. (*a*) If $\rho = .6$ show that the ratio between the major and minor axes of the ellipse is 2.

(*b*) Show that the slope of the regression line of y on x for a normal correlation surface is $\rho/(1 - \rho^2)^{1/2}$ in units of S_y and σ_x.

6. Establish the truth or falsity of the following proposition: A necessary and sufficient condition that two variables be normally correlated is that their regression systems be linear.

7. Prove that the regression systems of two normally correlated variables are linear and homoscedastic.

8. For (23) prove the following:
(*a*) the mean value of $\bar{y}_x$ taken over all values of x is zero,
(*b*) the variance of $\bar{y}_x$ is equal to $\rho^2\sigma_y^2$,
(*c*) the correlation coefficient between $\bar{y}_x$ and y is equal to ρ.

Hints. (*a*) Evaluate $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \bar{y}_x f(x, y)\, dy\, dx$,

(*b*) $$\sigma_{\bar{y}_x}^{\ 2} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \bar{y}_x^{\,2} f(x, y)\, dy\, dx,$$

(*c*) Evaluate $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{\bar{y}_x}{\sigma_{\bar{y}_x}} \frac{y}{\sigma_y} f(x, y)\, dy\, dx$.

9. If x and y are discrete variables, ρ is defined by

$$\rho = \frac{E(xy) - E(x)E(y)}{\sigma_x\sigma_y},$$

where

$$\sigma_x = [E(x^2) - \{E(x)\}^2]^{1/2}, \qquad \sigma_y = [E(y^2) - \{E(y)\}^2]^{1/2},$$

and

$$E(x) = \sum_1^n x_i g(x_i),$$

$$E(y) = \sum_1^m y_j h(y_j),$$

$$E(xy) = \sum_1^n\sum_1^m x_i y_j f(x_i, y_j),$$

$f(x_i, y_j)$ being the probability for the simultaneous occurrence of the pair of values (x_i, y_j), $g(x_i) = \sum_{j=1}^{m} f(x_i, y_j)$ and $h(y_j) = \sum_{i=1}^{n} f(x_i, y_j)$ being the marginal distributions of x and y, respectively. Find ρ for the table in Exercise 8, § 13, Chapter I.

10. Investigate the references given for tetrachoric r and give a report on the results of your study.

CHAPTER V

MULTIPLE AND PARTIAL CORRELATION

1. Notation. Simple correlation theory deals with co-variation in two variables. If other factors are involved the two variables are assessed as the important ones for the investigation and the other factors are ignored. But situations frequently arise in the fields of agriculture, biology, economics, education, and psychology, which call for consideration of three or more influences bearing simultaneously on a problem, and hence for the investigation of interrelations among three or more variables. For example, crop yield varies with soil fertility, rainfall, and temperature; wheat production is affected by acreage planted and yield per acre; students' honor points are connected with intelligence, health, hours of study, etc.; their chest measurements vary with stature and weight.

The term *multiple correlation* refers to a theory of correlation involving three or more variables. For ease in exposition we shall restrict the derivation of formulas to the three-variable case although

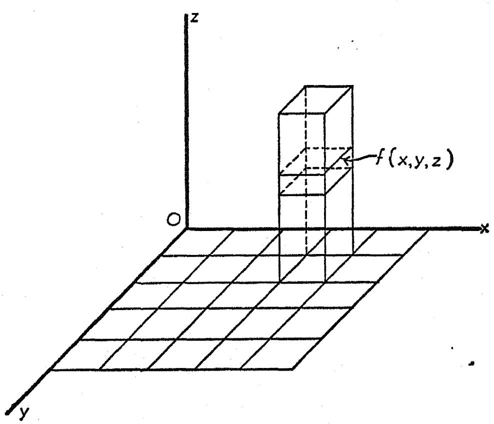


FIG. 12

the method is perfectly general. When the three-variable case is understood the formulas can be generalized for k variables.

The framework of a two-way table was a rectangle in the xy-plane

which was divided into cells by lines parallel to the axes. The analogue in the case of three variables, which we shall denote by x, y, and z, is a rectangular parallelopiped divided into cells by slicing planes parallel to the axes.

We shall denote the frequency in the cell whose mid-point has the coördinates (x, y, z) by $f(x, y, z)$. A pair of (x, y) values fixes a z column (Figure 12), and the sum of the frequencies in such a column is the " column total ":

$$\sum_z f(x, y, z) = f(x, y), \tag{1}$$

where here and subsequently the symbol $\sum$ together with the variable underneath denotes a summation in the direction of that variable. Now consider all those columns which have the same y. Their total frequency, denoted by

$$\sum_x f(x, y) = f(y), \tag{2}$$

may appropriately be called a " slab total " (Figure 13).

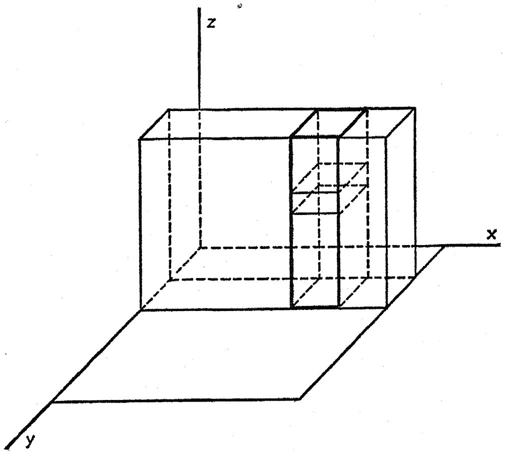


FIG. 13

Finally, if we add all the slab totals we get the total frequency N. Thus

$$\sum_y f(y) = N. \tag{3}$$

By making use of (1) we may, if we wish, express (2) as the double sum

(4) $$\sum_x \sum_z f(x, y, z) = f(y),$$

and using (4) we may express (3) as the triple sum

(5) $$\sum_x \sum_y \sum_z f(x, y, z) = N.$$

(a) The aggregate of the column totals $f(x, y)$ forms a two-way frequency table. If we imagine the numerical values of these frequencies written in the cells of the xy-plane it is easy to see that they constitute a correlation table (Figure 14). For this table, the simple correlation coefficient r_{xy} is called the total correlation (in contradistinction to a partial correlation coefficient to be defined later) and the regression curves are called the total regressions of y on x and x on y. Discussions analogous to (a) will now be given for horizontal columns parallel (b) to Ox and (c) to Oy.

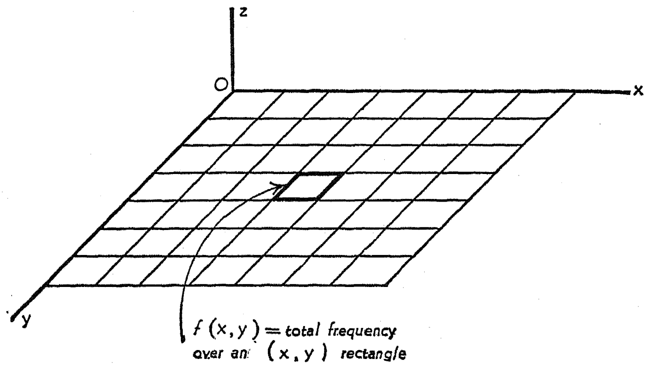


Fig. 14

(b) A pair of (y, z) values fixes an x column parallel to Ox. The sum of the frequencies in an x column is

(6) $$\sum_x f(x, y, z) = f(y, z).$$

If we add all those columns which have the same z we get a slab perpendicular to z whose total is

(7) $$\sum_y f(y, z) = f(z).$$

Finally, the totals of all such slabs is

(8) $$\sum_z f(z) = N.$$

The numerical values of the totals $f(y, z)$ written, if desired, in the cells of yz-plane form a two-way correlation table, as represented in Figure 15. For this table, r_{yz} is the total correlation coefficient between y and z, and the regression curves are the total regressions of y on z and z on y.

(c) Similarly, a pair of (x, z) values fixes a y column parallel to Oy. The sum of the frequencies in such a column is

$$(9) \qquad \sum_y f(x, y, z) = f(x, z).$$

If we add all the columns which have the same x we get a slab perpendicular to x whose total is

$$(10) \qquad \sum_z f(x, z) = f(x).$$

The sum of all such slabs is

$$(11) \qquad \sum_x f(x) = N.$$

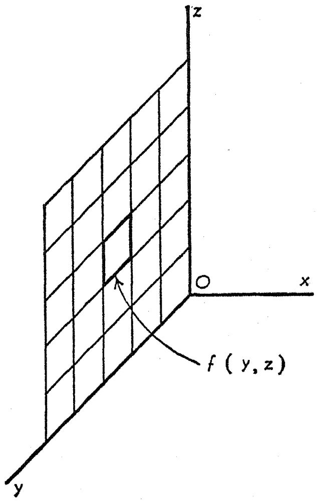


FIG. 15

The numerical values of the column totals $f(x, z)$ constitute a two-way correlation table whose correlation coefficient r_{xz} is the total

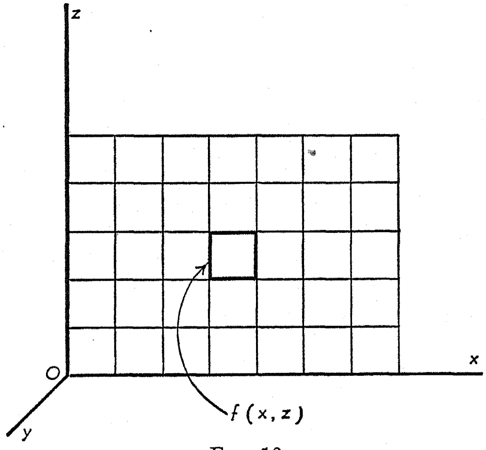


FIG. 16

correlation between x and z. The total regressions of x on z and z on x are given by the regression curves of this table (Figure 16).

2. Regression. The mean of a column at (x, y) is defined by

(12) $$\bar{z}(x, y) = \frac{1}{f(x, y)} \sum_z z f(x, y, z).$$

Similarly, the mean of an x column at (y, z) is

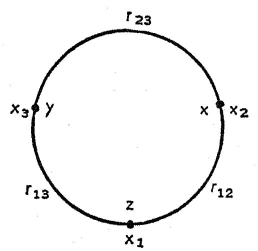
(13) $$\bar{x}(y, z) = \frac{1}{f(y, z)} \sum_x x f(x, y, z),$$

and the mean of a y column at (x, z) is

(14) $$\bar{y}(x, z) = \frac{1}{f(x, z)} \sum_y y f(x, y, z).$$

The *regression plane* of z on xy is that plane which fits the means of the z columns best in a least-squares sense. This should not be confused with the true regression surface, z on xy, which is defined as the locus of the mean points of the z columns. More accurately, it is the locus of these points as the dimensions of the cells approach zero. The regression plane, z on xy, is that plane which fits best the true regression surface, z on xy. Corresponding statements hold for the regression planes of y on xz and of x on yz.

So far, it was convenient to designate our variables by the conventional letters used in representing three-dimensional space. We are now about to obtain the equations of the regression planes and in order to extend our results to k variables it will be desirable to change to a new set of symbols which will lend themselves more readily to generalization. The switch will cause no difficulty. We shall now use x_1 in place of z, x_2 in place of x, and x_3 in place of y. The relations between the r's in the old notation and the new are $r_{xy} = r_{23}$, $r_{yz} = r_{13}$, $r_{xz} = r_{12}$. The adjacent diagram will help us keep in mind the relations between the new symbols and the old.



We shall now derive the equation of the regression plane of x_1 on x_2 and x_3. In determining, under a least-squares criterion, the parameters in its equation it will simplify the exposition if we assume that the variables are measured from their respective means as origin. This may be assumed without loss of generality. Let the desired equation be of the form

(15) $$x_1 = Ax_2 + Bx_3 + C.$$

Then we may determine the parameters in (15) so that the sum of the squares of the residuals

$$U = \sum_{1,2,3}(x_1 - Ax_2 - Bx_3 - C)^2 f \tag{16}$$

is a minimum, f being short for $f(x_1, x_2, x_3)$, and $\sum\limits_{1,2,3}$ for $\sum\limits_{x_1}\sum\limits_{x_2}\sum\limits_{x_3}$. Equating to zero the first partial derivatives of U with respect to A, B, and C, we obtain the equations

$$\begin{aligned} \sum x_2(x_1 - Ax_2 - Bx_3 - C)f &= 0, \\ \sum x_3(x_1 - Ax_2 - Bx_3 - C)f &= 0, \\ C &= 0. \end{aligned}$$

The simplification of the last equation is a consequence of our choice of origin since $\sum x_1 f = \sum x_2 f = \sum x_3 f = 0$ when the origin of x_i is at the mean of its N values. The first two equations may be written in the form

$$\begin{cases} A\sum x_2^2 f + B\sum x_2 x_3 f = \sum x_1 x_2 f, \\ A\sum x_2 x_3 f + B\sum x_3^2 f = \sum x_1 x_3 f. \end{cases} \tag{17}$$

Let σ_i^2 be the variance of x_i and let r_{ij} be the correlation coefficient between x_i and x_j. Then by definition,

$$\begin{aligned} \sum x_i^2 f(x_1, x_2, x_3) &= N\sigma_i^2, \\ \sum x_i x_j f(x_1, x_2, x_3) &= N\sigma_i\sigma_j r_{ij}. \end{aligned}$$

So (17) becomes

$$\begin{cases} NA\sigma_2^2 + NB\sigma_2\sigma_3 r_{23} = N\sigma_1\sigma_2 r_{12}, \\ NA\sigma_2\sigma_3 r_{23} + NB\sigma_3^2 = N\sigma_1\sigma_3 r_{13}. \end{cases} \tag{18}$$

Solving for A and B we have

$$A = \frac{\sigma_1}{\sigma_2}\,\frac{\begin{vmatrix} r_{12} & r_{23} \\ r_{13} & 1 \end{vmatrix}}{\begin{vmatrix} 1 & r_{23} \\ r_{23} & 1 \end{vmatrix}},$$

$$B = \frac{\sigma_1}{\sigma_3}\,\frac{\begin{vmatrix} 1 & r_{12} \\ r_{23} & r_{13} \end{vmatrix}}{\begin{vmatrix} 1 & r_{23} \\ r_{23} & 1 \end{vmatrix}}.$$

It is convenient both for simplicity and for the purpose of generalizing to k variables to define the determinant R by

$$R = \begin{vmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{vmatrix},$$

and to let R_{ij} be the cofactor of r_{ij}, that is, the minor of r_{ij} including the sign factor $(-1)^{i+j}$. Thus,

$$R_{12} = -\begin{vmatrix} r_{21} & r_{23} \\ r_{31} & r_{33} \end{vmatrix},$$

$$R_{13} = \begin{vmatrix} r_{21} & r_{22} \\ r_{31} & r_{32} \end{vmatrix}.$$

Clearly, $r_{11} = r_{22} = r_{33} = 1$, and $r_{12} = r_{21}$, etc., so the expressions for A and B may be written

$$A = -\frac{\sigma_1 R_{12}}{\sigma_2 R_{11}},$$

$$B = -\frac{\sigma_1 R_{13}}{\sigma_3 R_{11}}.$$

Hence (15) becomes

$$\frac{x_1}{\sigma_1} R_{11} + \frac{x_2}{\sigma_2} R_{12} + \frac{x_3}{\sigma_3} R_{13} = 0. \tag{19}$$

This equation gives the most probable value of x_1 for assigned values of x_2 and x_3, provided that the true regression is not far from being linear and the distribution of each x_1 column is nearly symmetrical so that its mode is close to its mean. It is an important equation because it shows how, on the average, changes in x_2 and x_3 affect x_1. The student will observe that the R's involve only simple correlation coefficients and that all the necessary computations for the terms in (19) were explained in Part I.

There are two analogous equations for the regression planes of x_2 on x_1x_3, and x_3 on x_1x_2, which can be obtained readily from (19) by a cyclical permutation of the subscripts on x and R. They are

$$\frac{x_2}{\sigma_2} R_{22} + \frac{x_3}{\sigma_3} R_{23} + \frac{x_1}{\sigma_1} R_{21} = 0 \tag{20}$$

when x_2 is the dependent variable, and

$$\text{(21)} \qquad \frac{x_3}{\sigma_3}R_{33} + \frac{x_1}{\sigma_1}R_{31} + \frac{x_2}{\sigma_2}R_{32} = 0$$

when x_3 is the dependent variable. Referred to an arbitrary origin (19) would have been

$$\text{(19a)} \qquad \frac{X_1 - \overline{X}_1}{\sigma_1}R_{11} + \frac{X_2 - \overline{X}_2}{\sigma_2}R_{12} + \frac{X_3 - \overline{X}_3}{\sigma_3}R_{13} = 0,$$

where $X_i - \overline{X}_i = x_i$. Analogous adjustments of (20) and (21) are obvious when the variables are referred to an arbitrary origin.

The three-dimensional case can now be generalized. By methods similar to those employed in deriving (15) we can derive the linear regression equation for k variables. Thus we have the hyperplane x_1 on $x_2, x_3, \cdots, x_k$,

$$\text{(22)} \qquad \frac{x_1}{\sigma_1}R_{11} + \frac{x_2}{\sigma_2}R_{12} + \cdots + \frac{x_k}{\sigma_k}R_{1k} = 0,$$

where R_{ij} is the cofactor of r_{ij} in

$$\text{(23)} \qquad R = \begin{vmatrix} r_{11} & \cdots\cdots & r_{1k} \\ \cdot\ \cdot\ r_{22} & \cdots\cdots & \cdot \\ \cdots\cdots & \cdots\cdots & \cdot \\ r_{k1} & \cdots\cdots & r_{kk} \end{vmatrix}.$$

When expressed in standard units, (22) becomes

$$\text{(22a)} \qquad t_1 = -\frac{1}{R_{11}}\sum_{i=2}^{k} R_{1i}t_i,$$

where $t_i = x_i/\sigma_i$. Then t_1 may be regarded as a weighted mean of the contributions of the other variables. The factor R_{1i} represents the force or weight of t_i when all these variables are given an opportunity to predict the value of t_1.

3. Standard Error of Estimate. In Part I (Chapter VIII) we learned that $S_y^2 = \sigma_y^2(1 - r^2)$ was a measure of the closeness with which the means of the x arrays of y clustered about the line of regression of y on x. S_y was called the standard error of estimate and the larger r was, the smaller was S_y. We now seek an analogous

expression for the three-variable case. To this end let

$$(24) \qquad S_{1.23}^2 = \frac{1}{N} \sum_{1,2,3} \delta^2 f(x_1, x_2, x_3)$$

where δ is the distance, measured parallel to the x_1-axis, between the regression plane and the points (x_1, x_2, x_3), $\sum\limits_{1,2,3}$ denoting a summation over all these points. That is, $\delta =$ (observed x_1 − estimated x_1), the estimated x_1 being given by (19). Then we may write

$$\begin{aligned}
NS_{1.23}^2 &= \frac{\sigma_1^2}{R_{11}^2} \sum \left(R_{11}\frac{x_1}{\sigma_1} + R_{12}\frac{x_2}{\sigma_2} + R_{13}\frac{x_3}{\sigma_3}\right)^2 f \\
&= \frac{\sigma_1^2 N}{R_{11}^2} (R_{11}^2 + R_{12}^2 + R_{13}^2 + 2R_{11}R_{12}r_{12} \\
&\qquad + 2R_{11}R_{13}r_{13} + 2R_{12}R_{13}r_{23}) \\
&= \frac{\sigma_1^2 N}{R_{11}^2} \{R_{11}(R_{11} + r_{12}R_{12} + r_{13}R_{13}) + R_{12}(R_{12} + r_{12}R_{11} + r_{23}R_{13}) \\
&\qquad + R_{13}(R_{13} + r_{13}R_{11} + r_{23}R_{12})\}.
\end{aligned}$$

According to Laplace's development of a determinant, the elements of any row (or column) and their corresponding cofactors may be used to develop R. If, in the resulting expression, the elements of this row (or column) are replaced by the corresponding elements of some other row (or column) the expression vanishes. Therefore, we have

$$(25) \qquad \begin{cases} R_{11} + r_{12}R_{12} + r_{13}R_{13} = R, & (a) \\ R_{12} + r_{12}R_{11} + r_{23}R_{13} = 0, & (b) \\ R_{13} + r_{13}R_{11} + r_{23}R_{12} = 0. & (c) \end{cases}$$

Using (25) in the above derivation we obtain

$$(26) \qquad S_{1.23}^2 = \frac{\sigma_1^2 R}{R_{11}} \cdot$$

This is a kind of *average variance* in x_1 columns of the observed values of x_1 from its corresponding estimated values on the regression plane (19). The square root of (26),

$$(26a) \qquad S_{1.23} = \sigma_1 \left(\frac{R}{R_{11}}\right)^{1/2}$$

is called the *standard error* of estimating x_1 from assigned values of x_2 and x_3.

4. Standard Deviation of Estimated Values. Next, we shall obtain an expression analogous to σ_{Ey} of Part I (§ 7, Chapter VIII) for the standard deviation of the estimated values given by (19). The mean value of these estimates is zero since x_i is measured from its mean as origin. Therefore, the variance, σ_{E1}^2, of the estimated values of x_1 is given by

$$
\begin{aligned}
(27) \qquad \sigma_{E1}^2 &= \frac{1}{N}\sum\left(\frac{\sigma_1}{\sigma_2}\frac{R_{12}}{R_{11}}x_2 + \frac{\sigma_1}{\sigma_3}\frac{R_{13}}{R_{11}}x_3\right)^2 f \\
&= \frac{\sigma_1^2}{R_{11}^2}(R_{12}^2 + R_{13}^2 + 2R_{12}R_{13}r_{23}) \\
&= \frac{\sigma_1^2}{R_{11}^2}\{R_{12}(R_{12} + R_{13}r_{23}) + R_{13}(R_{13} + R_{12}r_{23})\} \\
&= \frac{\sigma_1^2}{R_{11}^2}\{-R_{12}R_{11}r_{12} - R_{13}R_{11}r_{13}\} \text{ by } (b) \text{ and } (c) \text{ of } (25) \\
&= \frac{\sigma_1^2}{R_{11}}(R_{11} - R) \text{ by } (a) \text{ of } (25) \\
&= \sigma_1^2\left(1 - \frac{R}{R_{11}}\right).
\end{aligned}
$$

Hence we have

$$
(28) \qquad \sigma_{E1} = \sigma_1\left(1 - \frac{R}{R_{11}}\right)^{1/2}.
$$

If this result is to correspond to $\sigma_{Ey} = \sigma_y r$ we would expect that the factor $(1 - R/R_{11})^{1/2}$ would correspond in some way with r. This is indeed the case and we shall now show that this factor is the formula for the multiple correlation coefficient of x_1 on x_2 and x_3.

5. Multiple Correlation Coefficient. The ordinary correlation coefficient between the observed values of x_1 and its corresponding estimated values calculated from (19) is called the multiple correlation coefficient of x_1 on x_2 and x_3. It is denoted by $r_{1.23}$, so we have

$$
r_{1.23} = \frac{\sum {}_Ox_1\, {}_Ex_1}{N\sigma_1\sigma_{E1}},
$$

where ${}_Ox_1$ and ${}_Ex_1$ denote the observed and estimated values, respectively, of x_1.

Using (19) this may be written in the form

$$\begin{aligned} N\sigma_1\sigma_{E1}r_{1.23} &= \sigma_1^2\sum\frac{x_1}{\sigma_1}\left(-\frac{R_{12}}{R_{11}}\frac{x_2}{\sigma_2}-\frac{R_{13}}{R_{11}}\frac{x_3}{\sigma_3}\right) \\ &= \frac{N\sigma_1^2}{R_{11}}(-R_{12}r_{12}-R_{13}r_{13}) \\ &= \frac{N\sigma_1^2}{R_{11}}(R_{11}-R) \\ &= N\sigma_1^2\left(1-\frac{R}{R_{11}}\right). \end{aligned}$$

Making use of (28) in the above result we have the required formula

$$r_{1.23} = \left(1-\frac{R}{R_{11}}\right)^{1/2}. \tag{29}$$

By a cyclical permutation of the subscripts we can write at once the formulas for the multiple correlation coefficients of x_2 on x_1 and x_3, and of x_3 on x_1 and x_2. They are

$$r_{2.31} = \left(1-\frac{R}{R_{22}}\right)^{1/2}, \tag{30}$$

$$r_{3.12} = \left(1-\frac{R}{R_{33}}\right)^{1/2}. \tag{31}$$

By writing (26) in the form

$$S_{1.23}^2 = \sigma_1^2\left\{1-\left(1-\frac{R}{R_{11}}\right)\right\},$$

we obtain the formula

$$S_{1.23}^2 = \sigma_1^2(1-r_{1.23}^2) \tag{32}$$

which is quite analogous to the expression for S_y^2 in simple correlation. It is clear from (32) that

$$-1 \leqq r_{1.23} \leqq 1. \tag{33}$$

Each of the formulas (29), (30), and (31) may be generalized for k variables. Thus the multiple correlation coefficient of order $k-1$ of x_1 with the other $k-1$ variables is

$$r_{1.23\,\ldots\,k} = \left(1-\frac{R}{R_{11}}\right)^{1/2}, \tag{34}$$

where now R_{ij} is the cofactor of r_{ij} of R as defined in (23). While a mathematical generalization gives a more complete and aesthetic presentation, it is seldom that (22) or (34) are of value in practical cases for more than four variables.

For computing purposes it is pleasant to know that multiple correlation coefficients are expressible in terms of simple correlation coefficients.

Example 1. Three variables have in pairs simple correlation coefficients given by

$$r_{12} = .8, \qquad r_{13} = -.7, \qquad r_{23} = -.9.$$

Find the multiple correlation coefficient $r_{1.23}$ of x_1 on x_2 and x_3.

Solution.

$$R = \begin{vmatrix} 1 & .8 & -.7 \\ .8 & 1 & -.9 \\ -.7 & -.9 & 1 \end{vmatrix} = .068$$

$$R_{11} = .19, \qquad r_{1.23} = .8013.$$

Example 2. Suppose it is found that $r_{12} = .6$, $r_{13} = -.4$, $r_{23} = .7$. Comment on these results.

Solution. $R = -.346$, $R_{11} = .51$, $r_{1.23} = 1.29$. Inspecting the given r's we observe that large values of x_1 are associated with large values of x_2, but since r_{13} is negative it would mean that small values of x_1 go with large values of x_3 which is impossible when r_{12} and r_{23} are positive.

6. Limiting Cases. The following theorems are interesting in themselves and shed light on interpretations of the theory in applications.

Theorem I. *The necessary and sufficient condition for coincidence of the three regression planes* (19), (20), *and* (21), *is*

$$r_{12}^2 + r_{13}^2 + r_{23}^2 - 2r_{12}r_{13}r_{23} = 1. \tag{35}$$

Proof. From elementary analytic geometry, we know that a necessary and sufficient condition that two equations of the first degree represent the same plane is that their coefficients be proportional. For our equations this will be true when

$$\frac{R_{11}}{R_{12}} = \frac{R_{12}}{R_{22}} = \frac{R_{13}}{R_{23}}$$

and

$$\frac{R_{12}}{R_{13}} = \frac{R_{22}}{R_{23}} = \frac{R_{23}}{R_{33}}.$$

When expressed in terms of r_{ij} these relations, it will be found, all satisfy (35).

An *alternate proof* is as follows. When $S_{1.23}^2 = 0$ there is perfect functional dependence between the variables, assuming linear regression. It is evident from (26) that $S_{1.23}^2 = 0$ when $R = 0$. Upon expanding R in terms of r_{ij} and equating the result to zero we obtain (35).

COROLLARY. *Assuming linear regression, the criterion for perfect correlation between three variables is given by* (35).

Example 3. Given the following data, $r_{12} = .6$, $r_{13} = .4$. Find the value of r_{23} in order that $r_{1.23} = 1$.

Solution. Substituting the given values in (35) we have

$$r^2 - .48r - .48 = 0,$$

where the subscripts are dropped for the moment. Solving, we find $r = .24 \pm .73$. So $r_{23} = .97$.

The example shows that even though r_{12} and r_{13} are individually small, it does not follow that there cannot be high correlation between x_1, x_2, and x_3. Indeed two variables which individually with a third variable have correlations which are apparently worthless for predicting purposes may be very valuable when the three variables are taken together and multiple regression employed. On the other hand, it may be possible to get as good a prediction from r_{12} or r_{13} using simple regression as from multiple regression. This situation will be clarified by the following theorems.

Theorem II. *If* $r_{23} = 1$, *then* $r_{1.23}^2 = r_{12}^2 = r_{13}^2$, *and* $S_{1.23}^2 = \sigma_1^2(1 - r_{12}^2)$.

Proof. When $r_{23} = 1$ then $R_{11} = 0$ and it would appear from (29) that $r_{1.23}$ then becomes infinite. But this is impossible by (33). When $r_{23} = 1$ it will also happen that $r_{12} = r_{13}$. The student can easily verify this by letting $r_{23} = 1$ in (25) and subtracting (*c*) from (*b*) there. So we shall first see what (29) becomes when $r_{13} = r_{12}$. If in (*a*) of (25) we let $r_{13} = r_{12}$ we obtain $R - R_{11} = 2r_{12}R_{12}$, since R_{13} then equals R_{12}. Substituting this result in (29) we soon have

$$r_{1.23}^2 = \frac{-2r_{12}R_{12}}{R_{11}} = \frac{2r_{12}^2}{1 + r_{23}},$$

remembering that $r_{13} = r_{12}$. Now if we let $r_{23} = 1$ in the last expression we obtain the first conclusion of the theorem. The second conclusion follows from the first and formula (32).

In this case, then, multiple regression has no advantage over the simple regression x_1 on x_2 or x_1 on x_3, because the standard error is exactly what it would be if the third variable were not added. Since $r_{23} = 1$, there is perfect linear dependence between x_2 and x_3. Geometrically, all the data lie in the regression plane.

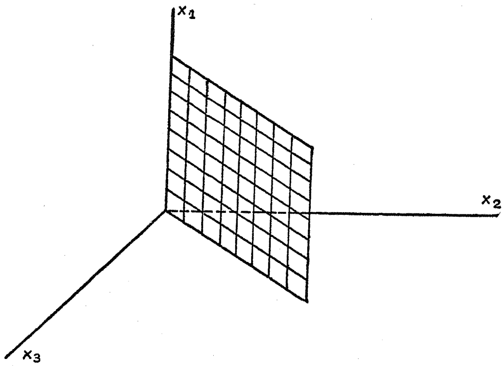


FIG. 17

Theorem III. *When* $r_{23} = 0$ *then* $r_{1.23}{}^2 = r_{12}{}^2 + r_{13}{}^2$.

Proof. When $r_{23} = 0$ it is easy to show that $R_{11} = 1$ and $R = 1 - r_{12}{}^2 - r_{13}{}^2$. So from (29) we have

$$r_{1.23}{}^2 = r_{12}{}^2 + r_{13}{}^2.$$

The formula for the standard error of prediction then becomes

$$S_{1.23}{}^2 = \sigma_1{}^2(1 - r_{12}{}^2 - r_{13}{}^2).$$

Hence, when x_2 and x_3 are completely independent, multiple regression gives a better prediction than would be given by either of the simple regressions x_1 on x_2 or x_1 on x_3; very much better if also r_{12} and r_{13} are nearly equal. If they are exactly equal their maximum value is $(\frac{1}{2})^{1/2} = .707$. This theorem shows that one has a good regression equation for predicting when each of two variables is highly correlated with the third variable but not with each other.

7. Partial Correlation. It is often important to measure correlation between two variables when the other variables have assigned values. For the case of three variables, to which we limit our attention, consider a slab parallel to the x_1x_2 plane (Figure 13). This is a sub-set of N which forms a two-way correlation table in which the relations between x_1 and x_2 hold for a fixed value of x_3. The

correlation coefficient between x_1 and x_2 in this sub-distribution is called the *partial correlation coefficient* between x_1 and x_2 for the assigned x_3 and is conventionally denoted by

$$r_{12.3}.$$

The regression curves for the table consisting of this sub-distribution are called the *partial regression curves.* A classical example of a partial correlation coefficient is the correlation between statures of fathers and sons when the stature of the mother is a particular value, say 62 inches.

In order to express $r_{12.3}$ in terms of the total correlations r_{ij}, as we were able to do in the case of $r_{1.23}$, it will be necessary to assume a theoretical or ideal situation. Suppose we are dealing with a distribution for which the total regression curves are straight lines and the regression surfaces are planes. Then the partial regression line, x_1 on x_2, in our table at x_3 will be a section of the regression plane, x_1 on x_2x_3, because the line will contain the mean points of all the x_1 columns, defined by the points (x_2, x_3), which lie in the table at x_3.

In the two variable case, described in Part I, we learned that S_y^2 was an average of the variances in the x arrays of y taken over all the values of x. Moreover, when the distribution was normal we proved that these variances were constant and S_y^2 was precisely this constant variance. The three variable case, in the ideal distribution we are about to consider, is quite analogous. Recall that $S_{1.23}^2$ could be regarded, in the ordinary case of linear regression, as an average of the variances of x_1 in the several columns at (x_2, x_3) since, when regression is linear, the means of the columns lie on the regression plane. Now let us assume that the distribution is homoscedastic in the x_1 direction so that the variances in all the columns of x_1 are the same. Under these assumptions, $S_{1.23}^2$ is the variance in *each* column of x_1's. Let $\sigma_{1.3}^2$ be the variance of x_1 in the table at x_3. Remember that $r_{12.3}$ is the correlation coefficient in this table and that regression is linear and homoscedastic. Therefore, for the variance $S_{1.23}^2$ in each of the columns of this table we may write

$$S_{1.23}^2 = \sigma_{1.3}^2(1 - r_{12.3}^2). \tag{36}$$

Now consider the two-way table of totals $f(x_1, x_3)$. In this table, r_{13} is the total correlation between x_1 and x_3, and $\sigma_{1.3}^2$ is the variance

in an x_3 array of x_1's. Since, under our assumption, $\sigma_{1.3}{}^2$ is constant over these arrays, we may write

$$(37) \qquad \sigma_{1.3}{}^2 = \sigma_1{}^2(1 - r_{13}{}^2).$$

From (32), (36), and (37) we obtain

$$(1 - r_{1.23}{}^2) = (1 - r_{13}{}^2)(1 - r_{12.3}{}^2)$$

that is

$$\frac{R}{R_{11}} = R_{22}(1 - r_{12.3}{}^2).$$

Solving, we have

$$r_{12.3}{}^2 = \frac{R_{11}R_{22} - R}{R_{11}R_{22}}.$$

By expanding the R's it is readily verified that

$$R_{11}R_{22} - R = (-R_{12})^2$$

is an identity. Therefore, we have the final result

$$(38) \qquad r_{12.3} = \frac{-R_{12}}{(R_{11}R_{22})^{1/2}}.$$

This may be written, if desired, in the form

$$(38a) \qquad r_{12.3} = \frac{r_{12} - r_{13}r_{23}}{\{(1 - r_{13}{}^2)(1 - r_{23}{}^2)\}^{1/2}}.$$

By letting $\sin\theta = r$, it is seen that tables of $\cos\theta = (1 - r^2)^{1/2}$ will facilitate the computation of (38a) in numerical problems.

Since (38a) does not involve x_3, the value of $r_{12.3}$ for one assignment of x_3 is the same as for any other assigned value of x_3. Therefore, not only must the distribution be homoscedastic in the x_1 direction, but also the value of $r_{12.3}$ in all slabs perpendicular to the x_3-axis must be the same. It is fairly obvious that these conditions would not, ordinarily, be satisfied in practical applications. So, in the applications, $r_{12.3}$ is regarded as a sort of average value of the partial correlations which could be obtained for all assignments of x_3. The chief use of partial correlation is in testing what the correlation between two variables would be if the third variable were not interfering with the relationship.

Example 4. In a study of the factors which influence "academic success," May* obtained the following results (among others) based on the records of 450 students at Syracuse University.

X_1 = honor points,	X_2 = general intelligence,	X_3 = hours of study,
$\overline{X}_1 = 18.5$,	$\overline{X}_2 = 100.6$,	$\overline{X}_3 = 24$,
$\sigma_1 = 11.2$,	$\sigma_2 = 15.8$,	$\sigma_3 = 6$,
$r_{12} = .60$,	$r_{13} = .32$,	$r_{23} = -.35$.

One purpose of the study was to find to what extent honor points were related to general intelligence, when hours of study (per week) are held constant. Using (38*a*) it is found that $r_{12.3} = .802$.

8. An Alternate Derivation. It is useful to approach the subject of partial correlation from another point of view. Assume, as before, that the variables x_1, x_2, x_3, are referred to the general mean as origin. Suppose that we wish to know what the correlation between x_1 and x_2 would be if the influence of x_3 were eliminated. Let us subtract from the x_1 of each point that part of x_1 which is due to the influence of x_3 as indicated by the regression line x_1 on x_3 and denote the residual by $x_{1.3}$. Then subtract from the x_2 of each point that part of x_2 which is due to x_3 as indicated by the regression line x_2 on x_3 and denote the residual by $x_{2.3}$. Thus we have

$$(39)\qquad \begin{cases} x_{1.3} = x_1 - r_{13}\dfrac{\sigma_1}{\sigma_3}x_3, \\[2ex] x_{2.3} = x_2 - r_{23}\dfrac{\sigma_2}{\sigma_3}x_3. \end{cases}$$

We shall now prove that the simple correlation coefficient between $x_{1.3}$ and $x_{2.3}$ is precisely $r_{12.3}$. By definition, this simple correlation coefficient would be

$$(40)\qquad \frac{\sum x_{1.3}x_{2.3}f(x_1, x_2, x_3)}{N\sigma_{1.3}\sigma_{2.3}}.$$

Making use of (39), the numerator of (40) becomes

$$\begin{aligned} \sum x_1x_2f - r_{13}\frac{\sigma_1}{\sigma_3}\sum x_2x_3f - r_{23}\frac{\sigma_2}{\sigma_3}\sum x_1x_3f + r_{13}r_{23}\frac{\sigma_1\sigma_2}{\sigma_3^2}\sum x_3^2f \\ = N(\sigma_1\sigma_2r_{12} - \sigma_1\sigma_2r_{13}r_{23} - \sigma_1\sigma_2r_{13}r_{23} + \sigma_1\sigma_2r_{13}r_{23}) \\ = N\sigma_1\sigma_2(r_{12} - r_{13}r_{23}). \end{aligned}$$

* *Predicting Academic Success* — Mark A. May, Journal Educational Psychology, 1923, vol. 14, 7, 429–440.

Now by (37),

$$\sigma_{1.3} = \sigma_1(1 - r_{13}^2)^{1/2}$$

and similarly,

$$\sigma_{2.3} = \sigma_2(1 - r_{23}^2)^{1/2}.$$

Inserting these results in (40) we obtain the promised result

$$\frac{r_{12} - r_{13}r_{23}}{\{(1 - r_{13}^2)(1 - r_{23}^2)\}^{1/2}}.$$

When interpreted according to this derivation, $r_{12.3}$ is sometimes called the " net " correlation between x_1 and x_2.

Interesting interpretations of multiple and partial correlation in terms of spherical trigonometry will be found in the following references:

1. Burgess, *The Mathematics of Statistics*, pp. 266–267; Houghton Mifflin Co.
2. Jackson, *The Trigonometry of Correlation*, Journal of the American Mathematical Association, vol. 31, pp. 275–280.

Exercises

1. Find the multiple correlation coefficients and the regression equations for the data in Example 4.

2. (*Garrett*) The r for intelligence and school achievement in a group of children 8 to 14 years old is .80. The r for intelligence and age in the same group is .70. The r for school achievement and age is .60. What will be the correlation between intelligence and school achievement in children of the same age?

3. (*Yule* and *Kendall*) The following means, standard deviations, and correlations are found for

X_1 = seed-hay crops in cwts. per acre,

X_2 = spring rainfall in inches,

X_3 = accumulated temperature above 42° F. in spring,

in a certain district in England during 20 years.

$\bar{X}_1 = 28.02$,	$\sigma_1 = 4.42$,	$r_{12} = .80$,
$\bar{X}_2 = 4.91$,	$\sigma_2 = 1.10$,	$r_{13} = -.40$,
$\bar{X}_3 = 594$,	$\sigma_3 = 85$,	$r_{23} = -.56$.

Find the partial correlations and the regression equation for hay crop on spring rainfall and accumulated temperature.

4. Derive and explain the relation $\sigma_1^2 = \sigma_{E1}^2 + S_{1.23}^2$. What is the corresponding relation in simple correlation?

5. The following data relate to land values and crops in twenty-five Iowa counties.

X_1 = average value per acre of farm land on January 1, 1920,
X_2 = average yield of corn per acre in bushels 1910–1919,
X_3 = per cent of farm land in small grain,
X_4 = per cent of farm land in corn.

County No.	X_1	X_2	X_3	X_4
1	$ 87	40	11	14
2	133	36	13	30
3	174	34	19	30
4	385	41	33	39
5	363	39	25	33
6	274	42	23	34
7	235	40	22	37
8	104	31	9	20
9	141	36	13	27
10	208	34	17	40
11	115	30	18	19
12	271	40	23	31
13	163	37	14	25
14	193	41	13	28
15	203	38	24	31
16	279	38	31	35
17	179	24	16	26
18	244	45	19	34
19	165	34	20	30
20	257	40	30	38
21	252	41	22	35
22	280	42	21	41
23	167	35	16	23
24	168	33	18	24
25	115	36	18	21

(*a*) Find the linear regression equation of X_1 on $X_2X_3X_4$.
(*b*) Estimate the first five values of X_1, using the equation obtained in (*a*).
(*c*) Calculate $S_{1.234}$ and $r_{1.234}$.

CHAPTER VI

FUNDAMENTALS OF SAMPLING THEORY WITH SPECIAL REFERENCE TO THE MEAN

1. Introduction.* To emphasize the viewpoint of the subject of this chapter it is convenient to recognize two general classes of problems in mathematical statistics. In problems of the first class our concern is largely with the exposition of methods of characterizing observed data. Thus in the first class would fall methods for summarizing the pertinent information in a set of variates by means of averages, measures of dispersion, indices of correlation, etc. In problems of the second class, however, the data at hand are regarded as a random sample drawn from a well-defined class of variates called the *population* or universe of discourse, and we are concerned with drawing inferences about the universe from the sample. By a *sample*, more precisely a *random sample*, we mean a sub-set of variates in which each individual from the universe has an equal and independent chance to be included. From this chosen sample we attempt to draw inferences concerning the universe. In order to deal with this inductive argument we first consider a deductive argument; that is, we first consider an infinite (or finite) universe and investigate the behavior of samples according to the laws of probability. The methodology dealing with this class of problems is known as sampling theory. Although the two classes of problems are not entirely distinct with regard to their treatment, the center of interest in sampling theory is the development of criteria for assisting common sense or educated judgment concerning the magnitude of chance fluctuations in statistical ratios, averages, and coefficients.

The Bernoulli theory deals with sampling fluctuations in relative frequencies. In the words of Professor Rietz,[1]

> But it is fairly obvious that the interest of the statistician in the effects of sampling fluctuations extends far beyond the fluctuations in relative frequencies. To illustrate, suppose we calculate any statistical measure such as an arithmetic

* A reference list is given at the end of each of the following chapters to which attention is directed in the course of the discussion by the use of superscripts.

mean, median, standard deviation, correlation coefficient, or parameter of a frequency function from the actual frequencies given by a sample of data. If we need then either to form a judgement as to the stability of such results from sample to sample or to use the results in drawing inferences about the sampled population, the common sense process of induction involved is much aided by a knowledge of the general order of magnitude of the sampling discrepancies which may reasonably be expected because of the limited size of the sample from which we have calculated our statistical measures.

A statistical measure calculated from the actual frequencies given by a sample has been called a *statistic* by R. A. Fisher.[2] This is to avoid a verbal confusion with the corresponding *parameter* in the universe which we should like to know but can generally only estimate. It is a matter of common experience that a statistic will vary from sample to sample. To characterize the variation that may be tolerated on the basis of chance is one of the fundamental problems of sampling theory.

In discussing such sampling fluctuations, Fisher[3] introduces the subject as follows:

> The idea of an infinite population distributed in a frequency distribution in respect of one or more characters is fundamental to all statistical work. From a limited experience, for example, of individuals of a species, or of the weather of a locality, we may obtain some idea of the infinite hypothetical population from which our sample is drawn, and so of the probable nature of future samples to which our conclusions are to be applied. If a second sample belies this expectation we infer that it is, in the language of statistics, drawn from a different population; that the treatment to which the second sample of organisms had been exposed did in fact make a material difference, or that the climate (or methods of measuring it) had materially altered. Critical tests of this kind may be called tests of significance, and when such tests are available we may discover whether a second sample is or is not significantly different from the first.

2. Method of Attack. The whole theory of sampling is based on frequency distributions and probability. In order to explain the tests of significance that have been developed, it is desirable to outline briefly the philosophy underlying the method of attack.

Sampling theory deals with specific questions like the following: Given the mean and standard deviation of a sample of N variates, how reliable are these estimates of the population mean and standard deviation, respectively? Given two samples, do their respective means or other statistics differ significantly? Can the differences be accounted for on the basis of chance or do the samples come from different populations? The answers require in general that we con-

ceive the universe as one distribution, the values of the statistic calculated from all possible samples of size N from that universe as another distribution, and that there are mathematical expressions capable of representing both distributions. This is the chief reason for studying frequency curves and probability distributions.

Suppose, for example, that we have computed a statistic — say the mean of 100 observations or measurements. What we get is not an absolutely fixed quantity which may be exactly reproduced again by taking 100 similar measurements. Indeed, if such an experiment were repeated many times, we would get values for the arithmetic mean which would form a frequency distribution. This distribution would have its own mean (mean of means), standard deviation, and higher moments. The law describing the frequency distribution of all possible means of samples of size N from a specified universe is called a distribution function when it can be expressed mathematically. Its graph is called the curve of means. What has been said of the mean holds similarly for any other statistic.

Formulation of statistical judgment about a sample involves the specification of the universe and the determination of the distribution function of a given statistic in samples of a given size drawn from this universe. The problem of determining the distribution functions for the various statistics from specified universes is one which has challenged modern mathematical research. In most cases it has been necessary to assume that the parent universe is of the normal form in order to obtain analytically the sampling distribution of the statistic. Many of the tests of significance are based upon this assumption. However, considerable information about sampling distributions from arbitrary universes is known in terms of their moments or expected values.

3. Expected Values. Let the continuous variable x be subject to the distribution function $f(x)$ and let $\phi(x)$ be an arbitrary function of x. Then the expected value of $\phi(x)$, denoted by application of the operator E, is defined by

$$E\{\phi(x)\} = \int_{-\infty}^{\infty} \phi(x)f(x)\,dx, \tag{1}$$

provided this integral exists. In particular, if $\phi(x) = x^k$, $(k = 1, 2, \cdots)$, we have

$$E(x^k) = \int_{-\infty}^{\infty} x^k f(x)\,dx.$$

For $k = 1$, this defines the mean of the x's in the universe represented by $f(x)$. Hereafter we will denote the mean of a universe of x's by $\tilde{x}$ and restrict $\bar{x}$ to denote the mean of a sample from that universe. Therefore, we may write *

$$E(x) = \tilde{x}. \tag{2}$$

If $\phi(x) = (x - \tilde{x})^2$, we have the variance of x,

$$\begin{aligned} \sigma_x^2 &= E(x - \tilde{x})^2 \\ &= E(x^2) - \tilde{x}^2. \end{aligned} \tag{3}$$

The (positive) square root of σ_x^2 is called the standard deviation or standard error of the distribution of x. Analogous definitions hold, of course, for y.

If the variables x and y are simultaneously distributed in accord with the function $f(x, y)$, then

$$E(xy) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf(x, y)\, dy\, dx.$$

If x and y are not independent variables in the probability sense, then, as we have seen in Chapter IV, $f(x, y) \neq g(x)h(y)$ where $g(x)$ and $h(y)$ are the marginal distributions of x and y, respectively. The correlation coefficient, ρ, between x and y in the bivariate universe represented by $f(x, y)$ is defined by

$$\rho = \frac{E(xy) - \tilde{x}\tilde{y}}{\sigma_x \sigma_y}. \tag{4}$$

The quantities $\tilde{x}$, σ, ρ, etc., relating to a universe are called parameters.

The following propositions may easily be established from preceding definitions so they are stated without proof.

I. *The expected value of the product of a variable and a constant is equal to the product of the constant and the expected value of the variable. That is,*

$$E(cx) = cE(x).$$

II. *The expected value of deviations of a variable from its expected value is zero. That is,*

$$E(x - \tilde{x}) = 0.$$

* $\tilde{x}$ is read "x tilde."

III. *The expected value of the sum of two or more variables is the sum of their expected values. In symbols,*

$$E(x + y + z) = E(x) + E(y) + E(z).$$

IV. *If x and y are mutually independent variables in the probability sense, then the expected value of their product is equal to the product of their expected values. That is,*

$$E(xy) = E(x)E(y).$$

V. *The expected value of the product of deviations of two mutually independent variables from their respective expected values is zero. That is,*

$$E\{(x - \bar{x})(y - \bar{y})\} = 0.$$

VI. *The expected value of the product of deviations of two correlated variables from their respective expected values is given by*

$$E\{(x - \bar{x})(y - \bar{y})\} = \rho\sigma_x\sigma_y.$$

4. Standard Error of a Linear Function of Variables. Suppose a variable is a linear function of two or more independent* variables each of which may take on a universe of values and we require the standard error of this function in terms of certain moments of the underlying distributions of independent variables. To this end let

$$w = c_1x_1 + c_2x_2 + \cdots + c_Nx_N \tag{5}$$

where each variable x_k, $(k = 1, 2, \cdots, N)$, is arbitrarily distributed and where the c's are arbitrary constants. Let σ_k represent the standard error of x_k in the universe to which it belongs, and let ρ_{ij} represent the correlation coefficient (if any correlation exists) between x_i and x_j. We seek the standard error of w, σ_w, in terms of σ_k and ρ_{ij}, $(i = 1 \text{ to } N, j = 1 \text{ to } N)$.

Case I. We will suppose first that the variables in the several universes are correlated, that is, that $\rho_{ij} \neq 0$ for every combination of i and j. From (5) and Proposition III we have

$$E(w) = c_1E(x_1) + c_2E(x_2) + \cdots + c_NE(x_N), \tag{6}$$

that is

$$\bar{w} = c_1\bar{x}_1 + c_2\bar{x}_2 + \cdots + c_N\bar{x}_N. \tag{7}$$

* We are using the phrase "independent variables" here in the ordinary sense of analysis to designate the variables on which a special function depends, without any implication that these variables are independent of each other in the statistical sense.

Then

$$E(w - \tilde{w})^2 = \sum c_i^2 E(x_i - \tilde{x}_i)^2 + \sum_{i \neq j} c_i c_j E(x_i - \tilde{x}_i)(x_j - \tilde{x}_j)$$

which by definition (3) and Proposition VI becomes

$$\sigma_w^2 = \sum c_i^2 \sigma_i^2 + \sum_{i \neq j} c_i c_j \rho_{ij} \sigma_i \sigma_j. \tag{8}$$

If $c_1 = 1$, $c_2 = \pm 1$, and $N = 2$, we have as a special case

$$\sigma_w^2 = \sigma_1^2 \pm 2\rho_{12}\sigma_1\sigma_2 + \sigma_2^2. \tag{9}$$

Case II. Suppose the x's in (5) are mutually *independent* in the statistical sense so that $\rho_{ij} = 0$. Then (8) becomes

$$\sigma_w^2 = c_1^2\sigma_1^2 + c_2^2\sigma_2^2 + \cdots + c_N^2\sigma_N^2. \tag{10}$$

5. Theorems. Relations (6)–(10) enable us to prove some interesting and useful theorems about the distribution of means of samples from an arbitrary universe. The following definition will make the notion of sample precise.

DEFINITION. *Let $(x_1, x_2, \cdots, x_N)$ be a set of N independent variables each subject to the same distribution function g, so that their joint distribution function is*

$$f(x_1, x_2, \cdots, x_N) \equiv g(x_1)g(x_2) \cdots g(x_N).$$

Then $(x_1, x_2, \cdots, x_N)$ is called a random sample of N from a universe with distribution function $g(x)$.

Table 6 exhibits the notation which will be used for the moments of the several distributions referred to in Theorems I–III.

TABLE 6. NOTATION

	Universe	*Sample*	*Distribution of Means*
Mean	$\tilde{x}$	$\bar{x}$	$E(\bar{x}) = \tilde{x}$
Standard Deviation	σ_x	s	$\sigma_{\bar{x}}$
Variance	σ_x^2	s^2	$\sigma_{\bar{x}}^2$
Skewness	$\tilde{\alpha}_{3:x}$	$\alpha_{3:x}$	$\tilde{\alpha}_{3:\bar{x}}$
Kurtosis	$\tilde{\alpha}_{4:x}$	$\alpha_{4:x}$	$\tilde{\alpha}_{4:\bar{x}}$

Theorem I. *If samples of size N be drawn from an arbitrary universe and if $\bar{x}$ be the mean of a sample, then the mean of all possible such means equals the mean of the universe. That is,*

$$E(\bar{x}) = \tilde{x}. \tag{11}$$

Proof. In (5), let $c_1 = c_2 = \cdots = c_N = 1/N$ and let $x_1, x_2, \cdots, x_N$, constitute a sample from a universe with mean $\tilde{x}$ and variance σ_x^2. Then $w = \bar{x}$. As a consequence of the definition of sample, $E(x_i) = \tilde{x}$ for each value of i from one to N. Therefore, (6) gives us $E(\bar{x}) = \tilde{x}$.

Theorem II. *The variance of the sampling distribution of means from an arbitrary universe equals the variance of the universe divided by the number in the samples. In symbols,*

$$\sigma_{\bar{x}}^2 = \frac{\sigma_x^2}{N}. \tag{12}$$

Hence,

$$\sigma_{\bar{x}} = \frac{\sigma_x}{(N)^{1/2}}. \tag{12a}$$

Proof. As in the proof of Theorem I let $w = \bar{x}$. Then (10) becomes

$$\sigma_{\bar{x}}^2 = \frac{1}{N^2}\sum_{1}^{N} \sigma_i^2. \tag{13}$$

Since the x's constitute a sample, $\sigma_i^2 = \sigma_x^2$ for each value of i from 1 to N. So (13) reduces to (12).

Theorem III. *The moments describing skewness and kurtosis in the sampling distribution of means are related to the corresponding moments in the universe by the following formulas:*

$$\begin{cases} \tilde{\alpha}_{3:\bar{x}} = \dfrac{\tilde{\alpha}_{3:x}}{\sqrt{N}}, \\ \tilde{\alpha}_{4:\bar{x}} = 3 + \dfrac{1}{N}(\tilde{\alpha}_{4:x} - 3). \end{cases} \tag{14}$$

A proof of (14) could be given by developing and applying additional propositions on expected values. However, this method is tedious for the higher moments. A more elegant proof can be given by means of characteristic functions.[4] Such a proof has been made available by Shewhart[5] for the discrete case.

The first and second theorems show us that in repeated samples, $\bar{x}$ is distributed about $\tilde{x}$ with standard deviation $\sigma_x/(N)^{1/2}$. Theorem III tells us something about the *form* of the distribution. Thus if the universe is normal so that $\tilde{\alpha}_{3:x} = 0$ and $\tilde{\alpha}_{4:x} = 3$, then from (14) we see that $\tilde{\alpha}_{3:\bar{x}} = 0$ and $\tilde{\alpha}_{4:\bar{x}} = 3$, so the sampling distribution of $\bar{x}$ from a normal universe has the normal values for skewness and kurtosis.

In the next three theorems it will be understood that x and y are correlated variables which are jointly distributed in accord with an arbitrary function $f(x, y)$ in which the parameters are $\tilde{x}$, $\tilde{y}$, σ_x, σ_y, and ρ.

Theorem IV. *Let $(x_1, y_1), (x_2, y_2), \cdots, (x_N, y_N)$, be a sample of N pairs drawn independently from the distribution characterized by $f(x, y)$ and let $(\bar{x}, \bar{y})$ be the mean of a sample. The correlation coefficient, R, between the means of all possible such samples equals ρ.*

Proof. By definition,

$$R = \frac{E(\bar{x}\bar{y}) - \tilde{x}\tilde{y}}{\sigma_{\bar{x}}\sigma_{\bar{y}}}, \tag{15}$$

and

$$\begin{aligned} E(\bar{x}\bar{y}) &= \frac{1}{N^2}E\{(x_1 + x_2 + \cdots + x_N)(y_1 + y_2 + \cdots + y_N)\} \\ &= \frac{E(S)}{N^2}, \end{aligned}$$

where

$$\begin{aligned} S = {} & x_1y_1 + x_1y_2 + \cdots + x_1y_N + \\ & x_2y_1 + x_2y_2 + \cdots + x_2y_N + \\ & \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad + \\ & x_Ny_1 + x_Ny_2 + \cdots + x_Ny_N. \end{aligned}$$

We will separate S into two parts, conveniently called u and v, where

$$u = x_1y_1 + x_2y_2 + \cdots + x_Ny_N,$$

and

$$v = \text{sum of } (N^2 - N) \text{ terms of the form } x_iy_j,\ i \neq j.$$

Then

$$E(u) = E\{\sum_1^N (x_iy_i)\} = \sum_1^N \{E(x_iy_i)\} = NE(xy).$$

In v, x_i must be uncorrelated with y_j since $i \neq j$. Therefore,

$$E(x_iy_j) = E(x_i)E(y_j) = \tilde{x}\tilde{y},$$

and

$$E(v) = (N^2 - N)\tilde{x}\tilde{y}.$$

So we have

$$E(S) = NE(xy) + (N^2 - N)\tilde{x}\tilde{y},$$

and therefore

$$E(\bar{x}\bar{y}) = \frac{1}{N}\{E(xy) + (N-1)\bar{x}\bar{y}\}. \tag{16}$$

Making use of Theorem II and (16) the right member of (15) reduces to the definition of ρ.

Theorem V. *Let $\bar{x}$ be the mean of a sample of N from $g(x)$ and let $\bar{y}$ be the mean of a sample of N from $h(y)$ where $g(x)$ and $h(y)$ are the marginal distributions of the universe characterized by $f(x, y)$ of correlated variables. Let $w = \bar{x} - \bar{y}$. The variance of the sampling distribution of w is*

$$\sigma_w^2 = \frac{1}{N}(\sigma_x^2 - 2\rho\sigma_x\sigma_y + \sigma_y^2). \tag{17}$$

The proof follows from (9) and Theorem IV.

Theorem VI. *Let $\bar{x}$ and $\bar{y}$ be the means, s_x and s_y the standard deviations, and r the correlation coefficient in a sample of N correlated items. Suppose N is so large that s^2 is a good estimate* of σ^2 and r of ρ, so that we may write*

$$\sigma_{\bar{x}}^2 = \frac{s_x^2}{N}, \qquad \sigma_{\bar{y}}^2 = \frac{s_y^2}{N}, \qquad \rho = r.$$

The variance of the sampling distribution of $w = \bar{x} - \bar{y}$ may be computed from the sample by the formula

$$\sigma_w^2 = \frac{1}{N^2}\left\{\sum(x_i - y_i)^2 - \frac{(\sum x_i - \sum y_i)^2}{N}\right\}. \tag{18}$$

The proof follows from (17).

6. An Experiment. We will now describe an exercise in experimental sampling which will help make the theory more meaningful. It was performed by a class of thirty students who took the distribution of Table 7 as a " universe."

In a box were placed 2000 discs† each bearing a number from the set 1, 2, 3, · · · , 25. The numbers on the discs were coded to the

* The problem of estimation is discussed in the next chapter.

† Small metal rimmed price tags were used. Ideally, each individual disc should be returned to the box before the next is drawn. However, this was not insisted upon and an entire sample may have been drawn before replacement.

TABLE 7. SPAN AMONG ADULT MALES. (See Table 20, Part I)

x	f
58.5	1
59.5	2
60.5	1
61.5	6
62.5	7
63.5	22
64.5	55
65.5	111
66.5	146
67.5	182
68.5	229
69.5	265
70.5	263
71.5	217
72.5	176
73.5	132
74.5	82
75.5	48
76.5	20
77.5	16
78.5	12
79.5	3
80.5	1
81.5	2
82.5	1

span values in accordance with the scheme shown on page 107, and the frequency of the variously numbered discs equaled the frequency of the corresponding x's. Each member of the class drew samples from the box according to the following directions.

Directions

1. Intermix the discs thoroughly and withdraw four random samples of ten discs each.
2. Record the numbers in each sample of ten on the sampling record sheet (page 107); replace the discs in the box.
3. For each sample of ten: find (*a*) mean span, (*b*) variance, (*c*) standard deviation.
4. Combine the four samples into a single sample of forty and find the statistics named in 3.

SAMPLING RECORD SHEET

Span	*Number on Disc*	*First Sample*	*Second Sample*	*Third Sample*	*Fourth Sample*	*Total*
58.5	1					
59.5	2					
60.5	3					
61.5	4					
62.5	5					
63.5	6					
64.5	7					
65.5	8					
66.5	9					
67.5	10					
68.5	11					
69.5	12					
70.5	13					
71.5	14					
72.5	15					
73.5	16					
74.5	17					
75.5	18					
76.5	19					
77.5	20					
78.5	21					
79.5	22					
80.5	23					
81.5	24					
82.5	25					
Mean *						
Standard Deviation						

* In computing the statistics let x denote span and u the number on a disc. Then $u = x - 57.5$, $\bar{x} = \bar{u} + 57.5$, and $s_x = s_u$.

The results of 3(*a*) will be reproduced here. There were, of course, 120 means from samples of $N = 10$. These were then grouped into a frequency distribution. The resulting distribution and its moments, together with the moments of the universe, are given in Table 8. (The computations were made according to the definitions given in Part I for the moments of an observed distribution.)

Although the chief purpose of the experiment is an appreciation of the theory, it is of interest to compare the experimental and

TABLE 8. DISTRIBUTION OF THE MEANS OF 120 SAMPLES OF N = 10 DRAWN FROM THE UNIVERSE OF SPAN

Interval	*Mid* $\bar{x}$	*Frequency*	*Moments*
67.0–67.3	67.15	1	Mean $\bar{\bar{x}} = 69.785$
67.4–67.7	67.55	1	$s_{\bar{x}} = 0.8941$
67.8–68.1	67.95	4	
68.2–68.5	68.35	4	$\alpha_{3:\bar{x}} = 0.052$
68.6–68.9	68.75	5	$\alpha_{4:\bar{x}} = 3.030$
69.0–69.3	69.15	19	
69.4–69.7	69.55	27	
69.8–70.1	69.95	20	$\tilde{x} = 69.943$
70.2–70.5	70.35	20	
70.6–70.9	70.75	7	$\sigma_x = 3.115$
71.0–71.3	71.15	6	$\tilde{\alpha}_{3:x} = 0.161$
71.4–71.7	71.55	3	
71.8–72.1	71.95	3	$\tilde{\alpha}_{4:x} = 3.296$

theoretical results. According to Theorem I the mean should be 69.943; we obtained 69.785. According to Theorem II the standard deviation should be $3.115/(10)^{1/2} \doteq .985$; we obtained .894. It is left as an exercise for the student to verify that the approximations of the α's are also close.

We may think of this "universe" as approximating a Type III curve and the distribution of Table 8 as approximating its sampling

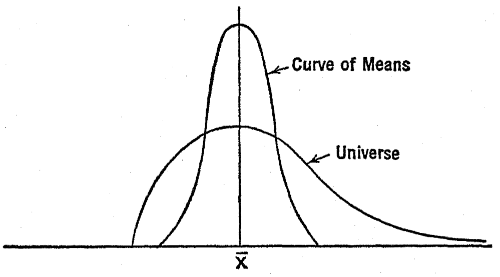


FIG. 18. DEPICTING THE SAMPLING DISTRIBUTION OF MEANS FROM A TYPE III UNIVERSE

curve of means (Figure 18). To represent graphically a universe and the curve of sample means from that universe would require analytical expressions for both these distributions. As yet, neither a type of universe has been specified nor has the functional form of the curve of means from that universe been determined. How-

ever, Figure 18 will help the student appreciate the meaning of some of the moment relations developed in § 5.

7. Reproductive Property of Normal Law. An important problem is to find the distribution function of the sum of several independent variables when these variables are normally distributed. It suffices to show how this problem can be solved for the sum of two such variables. The following discussion follows closely a proof given by Jackson.[6]

Let x and y be independent variables and normally distributed about zero as mean with standard deviations σ_1 and σ_2, respectively. Their distribution functions will have the forms

$$g(x) = C_1 e^{-ax^2}, \qquad h(y) = C_2 e^{-by^2}, \qquad a = \frac{1}{(2\sigma_1^2)}, \qquad b = \frac{1}{(2\sigma_2^2)};$$

the explicit values $1/C_1 = \sigma_1(2\pi)^{1/2}$, $1/C_2 = \sigma_2(2\pi)^{1/2}$, for total frequency 1, are not needed at the moment.

If $f(x, y)$ is the joint distribution function for x and y with marginal distributions $g(x)$ and $h(y)$ we shall first show that the frequency function, $H(w)$, for the variable $w = x + y$ is

$$H(w) = \int_{-\infty}^{\infty} f(x, w - x)\, dx.$$

For $\alpha < w < \beta$, when $\alpha - x < y < \beta - x$; these inequalities define a strip of the (x, y)-plane for which the corresponding frequency is

$$F(\alpha, \beta) = \int_{-\infty}^{\infty}\int_{\alpha-x}^{\beta-x} f(x, y)\, dy\, dx;$$

in the integration with respect to y, the substitution $w = x + y$, $y = w - x$, makes

$$\int_{\alpha-x}^{\beta-x} f(x, y)\, dy = \int_{\alpha}^{\beta} f(x, w - x)\, dw,$$

and hence

$$\begin{aligned} F(\alpha, \beta) &= \int_{-\infty}^{\infty}\int_{\alpha}^{\beta} f(x, w - x)\, dw\, dx \\ &= \int_{\alpha}^{\beta}\int_{-\infty}^{\infty} f(x, w - x)\, dx\, dw \\ &= \int_{\alpha}^{\beta} H(w)\, dw. \end{aligned}$$

We can now proceed with the main part of the proof. Since x and y are independent, their joint distribution may be written

$$\begin{aligned} f(x, y) &= g(x)h(y) \\ &= C_1C_2e^{-ax^2-by^2}, \end{aligned}$$

and so we have

$$\begin{aligned} H(w) &= \int_{-\infty}^{\infty} f(x, w - x)\, dx \\ &= C_1C_2\int_{-\infty}^{\infty} e^{-ax^2-b(w-x)^2}\, dx. \end{aligned}$$

To evaluate this integral write the exponential expression in the form

$$\begin{aligned} ax^2 + b(w - x)^2 &= (a + b)\left\{x - \frac{bw}{a + b}\right\}^2 + \frac{ab}{a + b}w^2 \\ &= (a + b)z^2 + cw^2, \end{aligned}$$

where

$$z = x - \frac{bw}{a + b}, \qquad c = \frac{ab}{a + b} = \frac{1}{2(\sigma_1^2 + \sigma_2^2)}.$$

The value of w being regarded as constant for the integration with respect to x, so that incidentally $dz = dx$, the expression for $H(w)$ can be written in the form

$$\begin{aligned} H(w) &= C_1C_2e^{-cw^2}\int_{-\infty}^{\infty} e^{-(a+b)z^2}\, dz \\ &= Ke^{-cw^2}, \end{aligned}$$

where

$$\begin{aligned} K &= C_1C_2\int_{-\infty}^{\infty} e^{-(a+b)z^2}\, dz \\ &= C_1C_2\left\{\frac{\pi}{a + b}\right\}^{1/2} \\ &= \frac{1}{\sigma_w(2\pi)^{1/2}}, \end{aligned}$$

and

$$\sigma_w^2 = \frac{1}{2c} = (\sigma_1^2 + \sigma_2^2).$$

If x, y, and u are independent and normally distributed, the quantity $x + y + u$ can be regarded as the sum of the two inde-

pendent normally distributed variables $x + y$ and u, and so is itself normally distributed. The conclusion can be carried over by induction, without further calculation, to the sum of any finite number of variables. Hence we have the following theorem.

Theorem VII. *If* $x_1, x_2, \cdots x_N$, *are independent variables and normally distributed with variances* $\sigma_1^2, \sigma_2^2, \cdots, \sigma_N^2$, *the function* $w = \sum_1^N c_i x_i$ *is normally distributed with variance* $\sigma_w^2 = \sum_1^N c_i^2 \sigma_i^2$.

The essential feature of the theorem is the part relating to the form of the distribution. This rather remarkable property of a linear function of normally distributed variables is sometimes called the *reproductive property* of the normal distribution. The part of the theorem relating to the magnitude of the variance follows necessarily from a general formula which was previously established without supposing the variables normally distributed or otherwise specialized.

Corollary. *The sampling distribution of means from a normal universe is itself normally distributed. The mean of the sampling distribution is the same as the mean of the universe and its variance is the variance of the universe divided by the size of the sample.*

The proof is left to the student. One should not conclude that it is generally true that the means of samples of N are distributed according to the same type of function which specifies the universe from which they are drawn. But the magnitudes of the mean and variance, as given in (11) and (12), are general in the sense that they are true for the sampling distribution of means from any infinite universe.

8. Non-Normal Universes. From analytic considerations, comparatively little is known at present about the exact distributions of statistics for samples drawn from non-normal universes. In a recent paper, Rietz[7] has listed the contributions and summarized the progress that has been made in this connection. The reader may refer to this paper.

With regard to the mean, Theorem III tells us that $\tilde{\alpha}_{3:\bar{x}} \to 0$ and $\tilde{\alpha}_{4:\bar{x}} \to 3$ as $N \to \infty$. So, even though the universe is far from normal, if the sample is made large enough, the sampling distribution of $\bar{x}$ approaches the normal form as characterized by skewness and kurtosis. (The conditions $\alpha_3 = 0$ and $\alpha_4 = 3$ are necessary but not sufficient conditions for a normal distribution.) Even for comparatively small values of N there is sufficient experimental evidence to

consider the distribution of $\bar{x}$ as normal to a high degree of approximation.

FINITE UNIVERSES. So far we have assumed that the universe was "infinite," that is, that it was indefinitely large in all its classes, as compared with the sample. This condition could be satisfied with a limited supply, for example in the experiment described in § 6, by replacement after each individual draw. However, if the entire sample is drawn from a limited supply before replacement, the probability of drawing an individual from a given class will be affected each time that one is drawn from that class. In such a case the universe is said to be "finite."

If M is the total frequency of a finite universe, the first four moments of the sampling distributions of $\bar{x}$ are as follows:

$$(19)\quad \begin{cases} E(\bar{x}) = \tilde{x} \qquad \sigma_{\bar{x}}^2 = \dfrac{M-N}{N(M-1)}\sigma_x^2, \\ \tilde{\alpha}_{3:\bar{x}}^2 = \dfrac{(M-1)(M-2N)}{N(M-N)(M-2)}\tilde{\alpha}_{3:x}^2, \\ \tilde{\alpha}_{4:\bar{x}} = \\ \dfrac{(M-1)\{(M^2-6MN+M+6N^2)\tilde{\alpha}_{4:x}+3M(M-N-1)(N-1)\}}{N(M-2)(M-3)(M-N)}. \end{cases}$$

Their origin is doubtful.[8] They are more general than the formulas given in (12) and (14) and reduce to them if $M \to \infty$.

The conclusion of investigators is that the distribution of means from nearly any finite universe is practically normal. In this connection the following striking example is given by Carver.[9]

A group of students chose arbitrarily the following most unusual distribution for a parent universe:

TABLE 9

x	f
15	9
3	2
29	43
405	189
1710	37
Total	280

and found the distribution of $\sum_{1}^{N} x_i = N\bar{x}$ of 1000 samples of twenty-five variates each shown in Table 10. It was obtained as follows.

TABLE 10

Class	*f*
5,000–	2
7,000–	54
9,000–	203
11,000–	310
13,000–	254
15,000–	130
17,000–	36
19,000–	9
21,000–	2
Total	1000

Two hundred and eighty Hollerith cards were punched with numbers corresponding to the two hundred and eighty variates of the parent population. The cards were thoroughly shuffled and then placed in a tabulating machine. After twenty-five cards had run through the electric tabulator their total was recorded. By repeating this procedure one thousand samples were readily obtained. It is thus possible to obtain experimentally some appreciation of the sensitivity of the sampling distribution of means to changes in population form. Carver concludes that if the sample N is fifty or larger and the population is at least ten times N, the parent population has relatively little control over the shape of the distribution of $\bar{x}$.

Another set of experiments was conducted by Shewhart[10] who comes to the following conclusion:

> Such evidence, supported by more rigorous analytical methods beyond the scope of the present discussion, leads us to believe that in almost all cases in practice we may establish sampling limits for averages of samples of four or more upon the basis of normal law theory.

9. Tchebycheff's Inequality. In (1) replace x by w, let $\phi(w) = (w - \tilde{w})^2$, and in the expression for $E\{(w - \tilde{w})^2\}$ replace all values of w larger than $\tilde{w} + \delta\sigma$ by $\tilde{w} + \delta\sigma$ and all values of w less than $\tilde{w} - \delta\sigma$ by $\tilde{w} - \delta\sigma$ where δ is a positive number. Then

$$E\{(w - \tilde{w})^2\} \geq k^2 + (\delta\sigma)^2 P_\delta \qquad (20)$$
$$\geq \delta^2\sigma^2 P_\delta,$$

where

$$k^2 = \int_{\tilde{w}-\delta\sigma}^{\tilde{w}+\delta\sigma} (w - \tilde{w})^2 f(w)\, dw \geqq 0,$$

and P_δ is the probability that w lies outside the interval $(\tilde{w} - \delta\sigma, \tilde{w} + \delta\sigma)$. From (20) we have

$$P_\delta \leqq \frac{1}{\delta^2}, \tag{21}$$

and therefore the following theorem.

Theorem VIII. *The probability is not more than $1/\delta^2$ that a value of w taken at random from the universe $f(w)$ will differ from its expected value by more than a multiple δ of its standard deviation.*

This theorem is known variously as Tchebycheff's theorem, criterion, or inequality. A striking property is its independence of the nature of the distribution of w. But the gain in generality must be paid for and the price is inadequate information about the particular. That is, the inequality (21) may be too wide to be of practical value in passing judgments on sampling fluctuations in a known or proposed distribution. Nevertheless, it does have some useful applications, two of which will now be given.

10. Law of Large Numbers. The Bernoulli theorem (Chapter I, § 7) can now be established. Let $w = x/s$, x being the number of successes in s trials. Then $\tilde{w} = p$. Let P_δ be the probability that x/s lies outside the interval $(p - \epsilon, p + \epsilon)$, where $\epsilon > 0$. We may take $\epsilon = \delta(pq/s)^{1/2}$, a multiple of the standard deviation of the relative frequency x/s. Accordingly, by Theorem VIII we have

$$P_\delta \leq \frac{1}{\delta^2}.$$

Since

$$\frac{1}{\delta} = \frac{(pq/s)^{1/2}}{\epsilon},$$

we obtain the inequality

$$P_\delta \leq \frac{p(1-p)}{s\epsilon^2}.$$

For any assigned ϵ, P_δ can be made arbitrarily small by increasing s. Thus x/s becomes increasingly reliable as an estimate of p as s increases.

The inequality of Tchebycheff can also be used to prove the stability of the means of large samples. Consider a sample of N from $f(x)$ in which the variance is σ^2. Let w be a linear function of the sample defined by

$$w = \frac{x_1 + x_2 + \cdots + x_N}{N}.$$

Suppose c^2 is a constant such that $\sigma^2 \leqq c^2$. Since $w = \bar{x}$, we have

$$\begin{aligned}\sigma_w{}^2 &= \frac{\sigma^2}{N} \\ &\leqq \frac{c^2}{N}.\end{aligned}$$

Let P be the probability that $(\bar{x} - \tilde{x})^2 > h^2$. That is, P is the probability that

$$\begin{aligned}(\bar{x} - \tilde{x})^2 &> \frac{Nh^2}{c^2} \cdot \frac{c^2}{N} \\ &> \frac{Nh^2}{c^2} \sigma_w{}^2.\end{aligned}$$

Therefore, from Theorem VIII,

$$P \leqq \frac{c^2}{Nh^2}.$$

Since c and h are fixed, P can be made arbitrarily small by taking N sufficiently large. Hence we have the following theorem.

Theorem IX. *The probability that the mean of a sample of N variates will differ numerically by more than a given positive number h from the mean of the universe can be made arbitrarily small by taking N sufficiently large.*

Under the conditions of the theorem, $\bar{x}$ is said to converge *stochastically* to $\tilde{x}$. This type of convergence, however, should not be confused with convergence in the sense of analysis.

11. Probability Scale of Sampling Fluctuations. Now that the *personae dramatis* have been assembled, we can state a theorem which tells us what the approximate probability is that the mean of a sample will deviate by an assigned amount from a hypothetical mean. We are assuming here that σ_x is known; the case where σ_x is unknown will be discussed later.

We know that $\bar{x}$ is (or tends to be) normally distributed about $\tilde{x}$ with standard deviation $\sigma_{\bar{x}} = \sigma_x/\sqrt{N}$. If the distribution of $\bar{x}$ be reduced to standard units by the transformation

$$t = \frac{\bar{x} - \tilde{x}}{\dfrac{\sigma_x}{\sqrt{N}}}, \tag{22}$$

then we know that t is approximately normally distributed about zero with standard deviation of unity. Hence we can refer to a normal probability scale for the probability that one would obtain a random sample for which $\bar{x}$ differs from $\tilde{x}$ by as much as $|\delta|$, where δ is expressed in the $\sigma_{\bar{x}}$ unit. So we have the following theorem.

Theorem X. *The probability Q_δ that a random sample from an infinite universe will have a mean, $\bar{x}$, which will be within an interval δ of the mean, $\tilde{x}$, of the universe is approximately*

$$Q_\delta = 2\int_0^\delta \phi(t)\,dt,$$

where δ is the observed value of t given by (22) and $\phi(t)$ is the normal curve. Then $P_\delta = 1 - Q_\delta$ is the approximate probability that $\bar{x}$ will not be within $|\delta|$ of $\tilde{x}$. If the universe is normal, P_δ gives the exact probability.

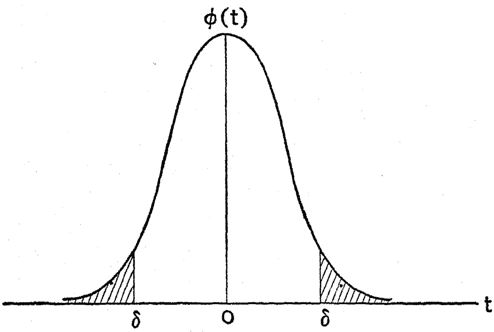


FIG. 19. $P_\delta =$. Q_δ IS THE PROBABILITY FOR A DEVIATION AS SMALL AS $|\delta|$, AND P_δ IS THE PROBABILITY FOR A DEVIATION AS LARGE AS $|\delta|$

12. Null Hypothesis and Significance Tests. The rationale underlying sampling theory has been summarized by E. S. Pearson[11] as follows:

In applying the methods of statistical analysis it is generally our aim to discriminate between two or more alternative hypotheses regarding the factors which have controlled certain observed events, which form what we term a sample or samples. If the process is examined in a little detail it will be found that the procedure may be described as follows:

(*a*) We define a hypothesis to be tested.

(*b*) We choose the criterion (or criteria) whose numerical value, derivable from the observations, is most suitable for testing the hypothesis. In doing this we recognize that the criterion is not a single-valued expression even if the hypothesis be true, but will vary from one sample of observations to another.

(*c*) We therefore refer the observed value of the criterion to this sampling distribution — *e.g.*, to a normal probability scale, etc. — and so obtain a measure of the likelihood of the hypothesis.

(*d*) Finally, if judged on this probability scale the observed criterion is not exceptional, we conclude that upon the information available there are no grounds for discarding the hypothesis; or if the value prove exceptional we consider the possibility of alternative hypotheses.

An hypothesis which is tested for possible rejection under the assumption that it is true has been called by Fisher[12] a *null hypothesis.* In other words, null hypothesis refers to a particular form of population distribution which is assumed in considering whether or not a sample could reasonably have arisen from the population which, in fact, was assumed. If the sample could not reasonably have arisen from the population proposed, as measured by a significance test, we say that the null hypothesis is refuted for the level of significance adopted. If the significance test yields a verdict of "not significant" for the probability level adopted, we say that the null hypothesis is not refuted or contradicted at that level.

It is open to the investigator to be more or less exacting concerning the smallness of the probability he would require before he would be willing to admit that his test has demonstrated a significant result. Good judgment in these matters comes only from much experience in the particular field in which the problem occurs. However, it is conventional among certain workers to adopt the following *rule:*

$$\begin{aligned}&\textit{If } P_\delta \geq .05,\ \delta \textit{ is not significant;}\\&\textit{if } P_\delta \leq .01,\ \delta \textit{ is significant;}\\&\textit{if } .05 > P_\delta > .01,\end{aligned}$$

our conclusions about δ *are doubtful and we cannot say with much certainty whether the deviation is significant or not until we have additional information.* Other workers prefer a more conservative level of significance.

Example 1. Suppose the mean span of 100 persons is found to be $\bar{x} = 70.56$ inches. Does this differ significantly from the mean $\tilde{x} = 69.943$ of the "universe" with standard deviation $\sigma_x = 3.115$? Calculating the above test we find $\delta = \dfrac{70.56 - 69.943}{3.115/\sqrt{100}} = 1.99$. Referring to the normal probability scale we find the chance of a difference between the observed and hypothetical means as large as that noted to be $P_\delta = .0471$. Our conclusion is that the given statistic $\bar{x} = 70.56$ is not exceptional, although it is possible that it came from a different universe, that is, in this case a different race of men.

Example 2. Twelve dice were thrown 26,306 times (Weldon's data), and a throw of 5 or 6 points was reckoned a success. The mean of the observed distribution was found to be 4.0524. In tossing a true die the chance of scoring 5 or 6 is $\frac{1}{3}$ so the number of dice scoring 5 or 6 should be distributed with frequencies proportional to the terms in the expansion $(\frac{2}{3} + \frac{1}{3})^{12}$. Therefore, the expected mean, on the hypothesis that the dice were true, is $sp = 12(\frac{1}{3}) = 4$. Test this hypothesis using the difference between the observed and theoretical means as a criterion of judgment.

Solution.

$$\sigma_x = (spq)^{1/2} = \left\{(12)\left(\tfrac{1}{3}\right)\left(\tfrac{2}{3}\right)\right\}^{1/2} = 1.633$$

$$N = 26{,}306,$$

$$\frac{\sigma_x}{N^{1/2}} = .010,$$

$$\delta = \frac{.0524}{.010} = 5.2.$$

The probability that a deviation outside $\delta = \pm 5$ would happen by chance is extremely small so we conclude that the dice were biased.

13. Size of Sample to Have a Given Reliability. From Theorem X we may determine the size N of a sample such that its mean, $\bar{x}$, will not differ from $\tilde{x}$ by more than a specified error $|\delta|$, with a degree of certainty equal to a specified probability.

Example 3. The American Rolling Mill Company investigated[13] the life of ferrous materials under different corrosive conditions. Data obtained from a certain kind of sheet material immersed in Washington tap water showed that the average time of failure of such sample was 874.89 days and the standard deviation of the time of failure was 85.31 days. There arose the following question of practical interest to the research engineer of this company: What sample size N must be used in order that for similar test conditions, the probability shall be 0.90 that the average time for failure determined from the N tests will be in error by not more than 5 per cent of the average of the universe?

Assuming that $874.89 = \tilde{x}$ and that means of samples of N are distributed normally, we may answer this question as follows: The allowable error is 5 per cent of 874.89 days or 43.74 days, and this must correspond to a probability of 0.90. From Theorem X we have

$$Q_\delta = 2\int_0^\delta = .90,$$

that is

$$\int_0^\delta = .45,$$

whence from the tables we find $\delta = 1.645$. Hence N is found by solving the equation

$$1.645 \frac{\sigma_x}{\sqrt{N}} = 43.74,$$

where $\sigma_x = 85.31$. We find $N = 10$.

14. Difference in Proportions. In the analysis of data obtained by sampling, certain problems occur which relate to the significance of apparent differences in proportions. Suppose we have two random samples of size n_1 and n_2, respectively, with x_1 individuals of the n_1 items and x_2 of the n_2 items which have a certain character or attribute. The question arises as to whether the observed difference is merely an accident of sampling or whether a similar difference exists in the universe. The following theorem may be used to test the null hypothesis that x_1/n_1 and x_2/n_2 are random and independent samples from the same universe.

Theorem XI. *If x_1/n_1 and x_2/n_2 are random and independent samples from an infinite universe in which p is the proportion of individuals which have the character in question, the probability that the difference in the proportions obtained will be numerically as great as the observed difference $w = |x_1/n_1 - x_2/n_2|$ is approximately P_δ, where P_δ is defined in Theorem X, and*

$$\delta = \frac{w}{\sigma_w}, \qquad \sigma_w = \left\{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right\}^{1/2}.$$

Proof. According to the Bernoulli theory, x_1/n_1 will vary about an expected value p with variance pq/n_1, where $q = 1 - p$. Similarly, x_2/n_2 will vary about p with variance pq/n_2. Then

$$E(w) = E\left(\frac{x_1}{n_1} - p\right) - E\left(\frac{x_2}{n_2} - p\right) = 0,$$

and from (10),

$$\sigma_w{}^2 = \frac{pq}{n_1} + \frac{pq}{n_2}. \tag{23}$$

Therefore, w varies about zero with variance given by (23), and the ratio

$$t = \frac{w}{\sigma_w} \tag{24}$$

varies about zero with unit standard deviation.

Information about the form of the t distribution may be obtained from its higher moments. It is not difficult to show that

$$(25) \qquad \begin{aligned} \tilde{\alpha}_3^2 &= \frac{1-4pq}{pq} \times \frac{(n_1-n_2)^2}{n_1n_2(n_1+n_2)}, \\ \tilde{\alpha}_4 &= 3 + \frac{1-6pq}{pq} \times \frac{n_1^2 - n_1n_2 + n_2^2}{n_1n_2(n_1+n_2)}. \end{aligned}$$

For fixed values of p and q, it is clear that $\tilde{\alpha}_3 \to 0$ and $\tilde{\alpha}_4 \to 3$ as the samples are taken indefinitely large. Even for moderately small samples the distribution of t does not differ greatly from the normal form. The following empirical rule, suggested by E. S. Pearson, is useful when one is in doubt about the propriety of referring (24) to the normal probability scale.

RULE. *Suppose $n_1 < n_2$ (we are at liberty to call either n_1). If $n_1p > 5$, the use of the normal probability scale is justified. If $n_1p \leq 5$, examine $\tilde{\alpha}_3^2$. If $\tilde{\alpha}_3^2 < .04$, it is still sufficiently accurate. But if $\tilde{\alpha}_3^2 \geq .04$, no great confidence can be placed in the test.*

In order to apply Theorem XI an estimate of p is usually required. For this purpose

$$(26) \qquad \hat{p} = \frac{x_1 + x_2}{n_1 + n_2}$$

is usually taken as the best estimate of p which is available from the samples. It is easy to show that $E(\hat{p}) = p$.

References

1. H. L. Rietz, *Mathematical Statistics*, p. 114. Open Court. 1927.
2. R. A. Fisher, *Foundations of Theoretical Statistics*. Phil. Trans. Royal Soc., vol. 222, (A), p. 309.
3. R. A. Fisher, *Statistical Methods for Research Workers*, § 11. Oliver and Boyd.
4. J. V. Uspensky, *Introduction to Mathematical Probability*. McGraw-Hill. 1937.
5. W. A. Shewhart, *Economic Control of Quality of Manufactured Product*, pp. 235–237. D. Van Nostrand. 1931.
6. Dunham Jackson, *The Theory of Small Samples*. American Mathematical Monthly, vol. 42, No. 6, pp. 344–364.
7. H. L. Rietz, *Some Topics in Sampling Theory*. Bulletin of the American Mathematical Society, April, 1937.
8. See Church, *Biometrika*, vol. 18, (1926), p. 357.
9. H. C. Carver, *Fundamentals of the Theory of Sampling*. Annals of Mathematical Statistics, vol. 1, No. 1, pp. 111–112.

10. *Loc. cit.*, p. 183.
11. E. S. Pearson, *The Test of Significance for the Correlation Coefficient.* Journal of the American Statistical Association, vol. 26, (1931), p. 128.
12. R. A. Fisher, *The Design of Experiments.* Oliver and Boyd. 1937.
13. See Shewhart, *loc. cit.*, p. 391.
14. B. H. Camp, *Mathematical Statistics*, p. 251, D. C. Heath and Co. 1931.

Problems

1. Suppose a variable w is normally distributed and a value is selected at random. Show that the odds are about 369 to 1 against the value differing from $E(w)$ by more than 3 σ_w's.
2. (*a*) Consider a finite universe of 5 variates: x_1, x_2, x_3, x_4, x_5. The number of distinct samples of 3 variates each that may be drawn is $C(5, 3) = 10$. Write these down.

   (*b*) Let $\bar{x}_i$ represent the ith sample mean and write down the 10 distinct sample means. For example,

   $$\bar{x}_1 = \frac{x_1 + x_2 + x_3}{3}.$$

   (*c*) Show that the mean of the 10 values of $\bar{x}_i$ is the mean of the 5 values of x_i. Thus,

   $$\frac{1}{10}\sum\bar{x}_i = \frac{1}{5}\sum x_i = \bar{x}.$$

   What formula does this example illustrate?
3. Show that the expected value of w^2 is greater than the square of the expected value of w.
4. From a box containing 2000 discs representing the distribution of span, draw a sample of 25 and compute its mean and standard deviation. Test the significance between your mean and the mean of the universe $\tilde{x} =$ 69.943 inches.
5. Suppose the weights of a sample of 1000 men of the same age are obtained yielding $\bar{x} = 140$ lbs. Assuming that $\sigma_x = 20.0$ lbs., what is the standard error of the mean of this sample? What is the probability that this mean does not differ from the mean of the universe at this age by more than five pounds?
6. (*Camp*[14]) The mean age of death of men who are alive at age 20 is, in the United States, 59.13. For the city of Chicago it is 58.98, and in 1910 the male population of age 20 was 24,000. Can the difference between the United States and Chicago be explained on the hypothesis of chance? Assume $\sigma_x = 10$ years, and that the distribution of the universe is approximately normal.
7. (*Camp*[14]) A fraternal organization wishes to be very sure that the average age of death in its group of men now aged 20 will not differ from the expected 59.13 years by more than one year. By " very sure " it means that Q_δ must equal .999 or more. How large should the group be? (Assume as before that $\sigma_x = 10$.)

8. Given that

$$w = \sum_1^k (f_i + x_i).$$

If the x's are independent and $\sum_1^k f_i$ is a constant, show that

$$\sigma_w{}^2 = \sum_1^k \sigma_i{}^2,$$

where $\sigma_i{}^2$ represents the variance of x_i.

9. Find the mean value of all positive ordinates of the first quadrant of $x^2 + y^2 = r^2$,

(*a*) when equally spaced along the x-axis,

(*b*) when equally spaced along the circle.

Answers:

(*a*) $$\frac{1}{r}\int_0^r y\,dx = \frac{1}{r}\int_0^r \sqrt{r^2 - x^2}\,dx = \frac{\pi r}{4},$$

(*b*) $$\frac{2}{\pi r}\int_0^r y\,ds = \frac{2}{\pi r}\int_0^r \sqrt{r^2 - x^2}\sqrt{1 + \left(\frac{dy}{dx}\right)^2}\,dx = \frac{2r}{\pi}.$$

10. Find the mean value of all the ordinates of the curve $y = a + b^x$ from 0 to x, when equally spaced along the x-axis.

11. Derive (25). *Hint.* $\tilde{\alpha}_r = E(t^r) = \dfrac{E(w^r)}{(\sigma_w)^r}$.

12. Show that the moment relations in (19) reduce to the corresponding relations in (12) and (14) if $M \to \infty$.

13. Suppose 300 mice having cancer of about the same degree of malignancy were divided at random into two groups of $n_1 = 100$ and $n_2 = 200$, respectively. The first group was given a certain serum treatment which was withheld from the second group but otherwise the two groups were treated alike. Among the serum treated there were $x_1 = 8$ deaths, and among the other group there were $x_2 = 25$ deaths. Test the significance of the difference between the mortality of 8% and $12\frac{1}{2}$% in the two groups.

14. An instructor had two classes of 20 and 30 students in the same subject. Four in the smaller class and 8 in the larger made grades of B or better. Should one seek a further explanation of this difference beyond variation due to sampling?

CHAPTER VII

SMALL OR EXACT SAMPLING THEORY

1. Introduction. A theory of sampling which assumes that N is large is inadequate for many practical problems. In recent years a theory has been developed to give more exact methods in dealing with small samples. In the practical field, the call for the solution of problems based on comparatively few observations was first realized in 1908 by a young man, then unknown, who chose to publish his results under the now celebrated pseudonym of "Student." Since then, many important contributions have been made toward the development and extension of this theory. Its applications are widespread. In the opinion of the present writer, continuity between large and small sample theory is an essential part of the newer attitude. In general, the methods of the theory of small sample theory are applicable to large samples, although the reverse is not true. It is our purpose in this chapter to facilitate an appreciation of some of the simpler aspects of this theory. The treatment centers around significance tests for means, variances, and correlation coefficients.

2. Expected Value of s^2. By definition, the variance of a sample is given by

$$s^2 = \frac{x_1^2 + x_2^2 + \cdots + x_N^2}{N} - \bar{x}^2. \tag{1}$$

Then the expected value of s^2 from repeated samples is

$$E(s^2) = E\left\{\frac{1}{N}(x_1^2 + x_2^2 + \cdots + x_N^2)\right\} - E(\bar{x}^2).$$

Since the x's constitute a sample we may write

$$E(x_1^2 + x_2^2 + \cdots + x_N^2) = NE(x^2),$$

and from (16) of Chapter VI, replacing y by x there, we have

$$E(\bar{x}^2) = \frac{1}{N}\{E(x^2) + (N-1)\tilde{x}^2\}.$$

Therefore,

$$E(s^2) = \frac{1}{N}\{NE(x^2)\} - \frac{1}{N}\{E(x^2) + (N-1)\bar{x}^2\}$$

$$= \frac{N-1}{N}\{E(x^2) - \bar{x}^2\}.$$

Hence

$$(2) \qquad E(s^2) = \frac{N-1}{N}\sigma^2$$

where σ^2 is the variance of x.

We may also obtain (2) as follows: Consider independent samples each containing N variates $u_1, u_2, \cdots, u_N$, where $u_i = x_i - \bar{x}$. For any sample,

$$s^2 = \frac{1}{N}\sum_1^N u_i^2 - \left\{\frac{1}{N}\sum_1^N u_i\right\}^2$$

$$= \frac{1}{N}\sum_1^N u_i^2 - \frac{1}{N^2}\sum_1^N u_i^2 - \frac{2}{N^2}\sum_1^N u_iu_j,\ i < j,$$

since the square of a sum is equal to the sum of the squares plus twice the cross-products. Then

$$E(s^2) = \frac{1}{N}E\{\sum_1^N u_i^2\} - \frac{1}{N^2}E\{\sum_1^N u_i^2\} - \frac{2}{N^2}E\{\sum_1^N u_iu_j\}.$$

By Proposition III of Chapter VI the right-hand member of the above expression may be written

$$\frac{1}{N}\sum_1^N\{E(u_i^2)\} - \frac{1}{N^2}\sum_1^N\{E(u_i^2)\} - \frac{2}{N^2}\sum_1^N\{E(u_iu_j)\},$$

which becomes

$$\frac{N\sigma^2}{N} - \frac{N\sigma^2}{N^2} - \frac{2}{N^2}\sum_1^N\{E(u_iu_j)\}.$$

Since $E(u_iu_j) = 0$, by Proposition V, we have the final result

$$E(s^2) = \frac{N-1}{N}\sigma^2.$$

This result is sometimes stated as in the following theorem.

Theorem I. *The mean of the sampling distribution of s^2 from an arbitrary universe equals the variance of the universe multiplied by the factor** $(N - 1)/N$.

It is to be anticipated that the expected value of s^2 is less than σ^2, as the following analysis will show. The variance σ^2 refers to deviations from $\tilde{x}$, whereas any s^2 refers to deviations from an $\bar{x}$. For any sample, then, we may regard $\tilde{x}$ as an arbitrary origin. Since in the case of any sample, the sum of the squares of deviations from its mean, $\bar{x}$, is less than the sum of the squares of deviations of the same variates from an arbitrary point $\tilde{x}$ (unless the sample is one whose mean falls at $\tilde{x}$), it is to be expected that the mean of all the values of s^2 will be less than σ^2. Relation (2) measures the extent of this inequality.

3. Unbiased Estimates of Population Parameters. A distribution function is not only a function of the variable involved, but it is also a function of the parameters, or hypothetical quantities, which are introduced to specify the universe sampled. In the case of a Bernoulli distribution the parameter is p, in the Poisson law it is m, and in a normal distribution there are two parameters, $\tilde{x}$ and σ.

A function of the variates given by a sample for estimating a parameter is called a *statistic.* Let $\hat{\theta}$ be a statistic corresponding to a parameter θ in the universe. We now state the following

DEFINITION. *If the expected value of $\hat{\theta}$, $E(\hat{\theta})$, equals θ then $\hat{\theta}$ is called an* ***unbiased estimate*** *of θ.*

It is clear from Theorem I of Chapter VI that the mean of a sample is an unbiased estimate of the mean of the universe. Also from (26) of Chapter VI we see that $\hat{p}$ defined there is an unbiased estimate of p.

Before the relation $\sigma_{\bar{x}}^2 = \sigma_x^2/N$ can be of much use to us in the applications we must have an estimate of σ_x^2 from the sample or samples available. By Proposition I of the preceding chapter,

$$E\left\{\frac{N}{N-1}s^2\right\} = \frac{N}{N-1}E(s^2)$$
$$= \sigma^2 \text{ by (2).}$$

* This factor is sometimes called "Bessel's correction." Perhaps it should be attributed more appropriately to Gauss who made use of it, in this connection, as early as 1823.

Let $\hat{\sigma}^2$ be an unbiased estimate of σ^2. If this estimate[1] is based on a single sample we have

$$(3) \qquad \hat{\sigma}^2 = \frac{N}{N-1}s^2 = \frac{\sum_1^N (x_i - \bar{x})^2}{N-1}.$$

If $n = N - 1$ it is obvious that

$$(3a) \qquad s^2 = \frac{n}{n+1}\hat{\sigma}^2.$$

It is conventional[2] to take

$$(4) \qquad \hat{\sigma} = \left\{\frac{N}{N-1}\right\}^{1/2} s$$

as an estimate of σ. If N is large the difference between unity and the coefficient of s in (4) is negligible in numerical problems. With N large it would not be invalid, to any appreciable extent, to use s as an estimate of σ.

If two independent samples are available from the same universe, an unbiased estimate based on the two samples is given by

$$(5) \qquad \hat{\sigma}^2 = \frac{q}{N-2},$$

where

$$q = N_1 s_1^2 + N_2 s_2^2, \; N = N_1 + N_2,$$

s_1^2 and s_2^2 being the variances of samples consisting of N_1 and N_2 variates, respectively. It is left as an exercise for the student to verify that the expected value of $q/(N - 2)$ is σ^2.

In case k independent samples are available from the same universe, we may generalize (5) and write

$$(6) \qquad \hat{\sigma}^2 = \frac{Q}{U-k},$$

where

$$Q = N_1 s_1^2 + N_2 s_2^2 + \cdots + N_k s_k^2,$$
$$U = N_1 + N_2 + \cdots + N_k,$$

and s_i^2 is the variance in the ith sample consisting of N_i variates.

When $\hat{\sigma}^2$ is used in future discussions it will be clear from the context whether this estimate is based on 1, 2, or k samples.

If $N_i = N$ is the same for every sample, (6) reduces to

$$\hat{\sigma}^2 = \frac{N(s_1^2 + s_2^2 + s_3^2 + \cdots + s_k^2)}{U - k}, \tag{7}$$

where $U = Nk$. Clearly, (7) may be written in the form

$$\frac{N-1}{N}\hat{\sigma}^2 = \frac{1}{k}(s_1^2 + s_2^2 + s_3^2 + \cdots + s_k^2). \tag{7a}$$

When k is taken infinitely large so that U becomes the universe, the right member of (7a) then refers to the expected value of s^2, and $\hat{\sigma}^2$ becomes σ^2 itself. So as $k \to \infty$ the limiting value of (7a) becomes

$$\frac{N-1}{N}\sigma^2 = E(s^2),$$

as given in (2).

As an alternate to (7), in the case where all samples contain the same number of variates, we may take

$$\begin{aligned} \hat{\sigma} &= \frac{1}{b(N)} \times \frac{1}{k}(s_1 + s_2 + s_3 + \cdots + s_k) \\ &= \frac{1}{b(N)} \times \text{mean value of standard deviations,} \end{aligned} \tag{8}$$

where $b(N)$ is a function of N and approaches unity as N increases. The exact expression for $b(N)$ will be derived in § 7. Its approximate value is $b(N) = 1 - 3/(4N)$. As $k \to \infty$ the limiting value of (8) becomes

$$\sigma = \frac{E(s)}{b(N)}. \tag{9}$$

In § 7 we will prove that $b(N)\sigma$ is the mean of the sampling distribution of s from a normal universe whose standard deviation is σ. Values of $b(N)$ and its reciprocal have been tabulated by E. S. Pearson[3] and others,[4] and we have included a short table in § 7.

As an alternate to (4) we have from (8) when $k = 1$,

$$\hat{\sigma} = \frac{s}{b(N)}. \tag{10}$$

4. Degrees of Freedom. In § 2 we have proved, essentially, that the expected value of $\sum(x_i - \bar{x})^2$ is $(N-1)\sigma^2$, where the N values of x in the sample are subject to the linear restriction $\sum x_i = N\bar{x}$. This is equivalent to proving that the expected value of $\sum x_i^2$ is $(N-1)\sigma^2$ when the x's are subject to the linear restriction $\sum x_i = 0$. Suppose, however, that there are $k < N$ linear restrictions on the x's. What, then, is the expected value of $\sum x_i^2$? A. T. Craig[5] has proved analytically that if $x_1, x_2, \cdots, x_N$, are N independent values of a variable which is normally distributed about zero with variance σ^2 and if the N values of x are subject to $k < N$ homogeneous linear restrictions, then the expected value of $\sum x_i^2$ is $(N-k)\sigma^2$. The number $n = N - k$ is frequently called the number of *degrees of freedom.*

5. "Student's" Distribution. The formula used in testing a null hypothesis that a given sample comes from a universe with a proposed mean is

$$t = \frac{(\bar{x} - \tilde{x})(N)^{1/2}}{\sigma}. \tag{11}$$

As stated in Chapter VI, (11) is normally distributed if the universe is normal. On the side of applications, σ is seldom available and usually must be estimated from the data available. If we substitute into (11) the estimate of σ given in (4) and calculate

$$t = \frac{(\bar{x} - \tilde{x})(N-1)^{1/2}}{s}, \tag{12}$$

we are not justified in asserting that (12) is normally distributed unless N is large. And so, in testing the significance of the mean of a small sample we are not justified in referring (12) to a normal probability scale. The variability of s from sample to sample invalidates that procedure.

While Helmert obtained the distribution of s^2 as early as 1876 it seems that "Student"[6] was the first to recognize the importance, for the theory of small samples, of taking account of the variability of s in (12). By means of a remarkable intuition he obtained, somewhat empirically, the joint distribution function of $\bar{x}$ and s from a normal universe. Later writers, notably Fisher, established his results rigorously.

"Student" actually found the distribution of a slightly different variable, *viz.*,

$$z = \frac{\bar{x} - \tilde{x}}{s}. \tag{13}$$

Obviously, z is functionally related to t by

(14) $$z = t(N - 1)^{-1/2},$$

so the distribution of t can easily be obtained from that of z. In deriving the distribution of z we shall follow the proof given by Fisher.[7] To avoid interrupting the main development some of the details will be deferred to the next section.

Consider a normal universe with frequency element

$$df = (2\pi\sigma^2)^{-1/2}e^{-(x-\bar{\bar{x}})^2/2\sigma^2}\,dx.$$

Let a sample $(x_1, x_2, \cdots, x_N)$ be taken at random from it. Then the probability that the sample will lie in the element of volume

$$dv = dx_1\,dx_2 \cdots dx_N$$

is

(15) $$dF = (2\pi\sigma^2)^{-N/2}e^{-V^2/2\sigma^2}\,dv,$$

where $V^2 = \sum_1^N (x_i - \bar{\bar{x}})^2$. From the relation* $\mu_2 = \nu_2 - \nu_1{}^2$ we have

$$V^2 = Ns^2 + N(\bar{x} - \bar{\bar{x}})^2.$$

Hence (15) may be written

(16) $$dF = (2\pi\sigma^2)^{-N/2}e^{-\{Ns^2+N(\bar{x}-\bar{\bar{x}})^2\}/2\sigma^2}\,dv.$$

By means of N-dimensional geometry (to be explained in § 6) Fisher showed that the element of volume dv can be expressed in terms of the variation of $\bar{x}$, namely, $d\bar{x}$, and the variation in volume, $d(s^{N-1})$, of an $(N - 1)$-dimensional hypersphere of radius $(N)^{1/2}s$, so that

(17) $$dv = Cs^{N-2}\,d\bar{x}\,ds,$$

where C is a constant. Then (16) becomes

(18) $$dF = ke^{-\{Ns^2+N(\bar{x}-\bar{\bar{x}})^2\}/2\sigma^2}s^{N-2}\,ds\,d\bar{x}.$$

From (18) the distribution of z can be deduced. From (13) we obtain $d\bar{x} = s\,dz$ for a fixed value of s. Substituting in (18) we obtain, for the joint distribution of s and z,

(19) $$ke^{-Ns^2(1+z^2)/2\sigma^2}s^{N-1}\,ds\,dz.$$

This expression is defined for $s \geq 0$, being identically zero for $s < 0$ since s is taken as the positive square root of s^2. If s is integrated

* Cf. Part I, Ch. IV, § 9.

out of (19), the distribution of the single variable z is obtained. To perform this integration, let

$$y = s(1 + z^2)^{1/2}, \qquad ds = (1 + z^2)^{-1/2}\, dy,$$

and integrating with respect to y from 0 to ∞, we have

$$k\left\{\int_0^\infty y^{N-1} e^{-Ny^2/2\sigma^2}\, dy\right\}(1 + z^2)^{-N/2}\, dz$$

which reduces to

$$K(1 + z^2)^{-N/2}\, dz$$

where, as shown in § 4 of Chapter II,

$$K = \frac{k}{2}\left(\frac{2\sigma^2}{N}\right)^{N/2}\Gamma\left(\frac{N}{2}\right) = \frac{\Gamma\left(\dfrac{N}{2}\right)}{(\pi)^{1/2}\Gamma\left(\dfrac{N-1}{2}\right)}. \tag{20}$$

Therefore, the distribution function for "Student's" z is

$$F(z) = \frac{\Gamma\left(\dfrac{N}{2}\right)}{(\pi)^{1/2}\Gamma\left(\dfrac{N-1}{2}\right)}(1 + z^2)^{-N/2}. \tag{21}$$

The curve is symmetrical with mean zero and infinite range. It is quite different, however, in mathematical character from the normal curve although it approaches this form as $N \to \infty$. (Cf. § 9.) From the viewpoint of sampling theory the important property of (21) is its independence of σ. The revolutionary character of this property is revealed in certain applications that involve drawing probable inferences from small samples, say from a sample of $N = 10$.

Using (14) Fisher modified (21) and obtained the distribution of t which is the one now widely applied. Before discussing the t-distribution, we shall give the details of Fisher's derivation of (21) and consider the separate distributions of $\bar{x}$, s^2, and s.

6. Fisher's Derivation. Making use of the geometrical method employed by Fisher[7] we shall imagine an N-dimensional space in which we take the origin at the point $O(\bar{x}, \bar{x}, \cdots, \bar{x})$ and rectangular axes Ou_1, Ou_2, $\cdots$, Ou_N. A point can be located in a space of a

specified number of dimensions by associating with the point a set of numbers. Therefore, we may represent the sample by the point $P(u_1, u_2, \cdots, u_N)$ where $u_i = x_i - \bar{x}$. Although it is impossible to visualize a space of N dimensions for $N > 3$ we will carry through the argument for the general case by analogy with the case for $N = 3$. So we consider the latter case first.

When $N = 3$, the sample is represented by the point $P(u_1, u_2, u_3)$ and we have the mean $\bar{u}$ and variance s^2 defined by

(a) $$u_1 + u_2 + u_3 = 3\bar{u}$$

and

(b) $$(u_1 - \bar{u})^2 + (u_2 - \bar{u})^2 + (u_3 - \bar{u})^2 = 3s^2.$$

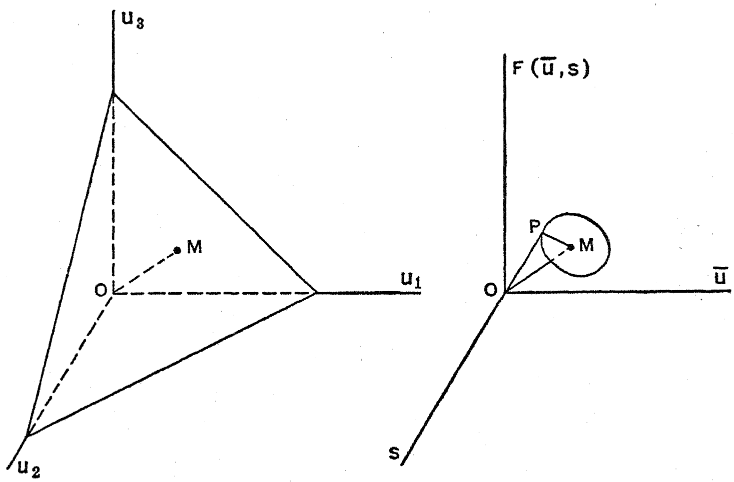


FIG. 20

For an assigned $\bar{u}$, (a) represents a plane; and, for an assigned pair of values of $(\bar{u}, s)$, (b) represents a sphere with center at the point $M(\bar{u}, \bar{u}, \bar{u})$. The line

(c) $$u_1 = u_2 = u_3$$

has direction cosines each equal to $1/(3)^{1/2}$ and is normal to the plane (a). The perpendicular distance of P from this line is

$$MP = s(3)^{1/2}$$

as can be seen from (b). We require the probability, to within infinitesimals of order higher than $d\bar{u}\, ds$, of getting a sample of

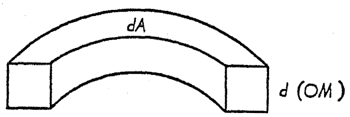
$N = 3$ independent values of u which will simultaneously yield values of $\bar{u}$ and s which lie within the region bounded by $\bar{u}$, $\bar{u} + d\bar{u}$ and s, $s + ds$. Following the method of § 5, an element of this probability density is given by the expressions

$$\begin{aligned} dF &= (2\pi\sigma^2)^{-3/2} e^{-(u_1^2+u_2^2+u_3^2)/2\sigma^2}\, dv \\ &= (2\pi\sigma^2)^{-3/2} e^{-3(s^2+\bar{u}^2)/2\sigma^2}\, dv \end{aligned}$$

where

$$dv = du_1\, du_2\, du_3.$$

As the sample point $P(u_1, u_2, u_3)$ varies, $\bar{u}$ and s also vary. Corresponding to different values of s there are a set of concentric spheres defined by (b) all having the same center. Since the plane (a) passes through the common center of the spheres, the region dv is a shell between concentric spheres of radii $\sqrt{3}s$ and $\sqrt{3}(s + ds)$. To use a homely illustration, dv corresponds to one of the successive layers in an onion. Our problem is to express dv in terms of $\bar{u}$, s, $d\bar{u}$, and ds. Now the line (c) meets the plane (a) at M and the distance OM is



$$OM = \bar{u}(3)^{1/2}$$

so we have the differential element

$$d(OM) = (3)^{1/2}\, d\bar{u}.$$

Since the plane (a) passes through M, the intersection of the plane and sphere is a great circle with center at M and radius equal to $s(3)^{1/2}$. The area of this circle is

$$A = 3\pi s^2$$

and the differential element dA is

$$dA = 6\pi s\, ds.$$

Therefore, within infinitesimals of higher order,

$$\begin{aligned} dv &= dA\, d(OM) \\ &= C_1 s\, ds\, d\bar{u} \end{aligned}$$

where here and hereafter, in this section, the C's are constants. So, the required probability is

$$dF = C_2 e^{-3(s^2+u^2)/2\sigma^2} s\, ds\, d\bar{u}.$$

Passing now to the general case involving N-space, let P be the

point representing the sample $(u_1, u_2, \cdots, u_N)$. Then PM is the perpendicular from P upon the line

$$(d) \qquad u_1 = u_2 = \cdots = u_N$$

and we have

$$OM = (N)^{1/2}\bar{u},\ \overline{OP}^2 = \sum u^2,$$
$$\overline{MP}^2 = \overline{OP}^2 - \overline{OM}^2 = \sum u^2 - N\bar{u}^2 = Ns^2.$$

In N-space, the plane (a) generalizes into the hyperplane

$$(e) \qquad \sum u_i = N\bar{u},$$

and the sphere (b) generalizes into the hypersphere

$$(f) \qquad \sum (u_i - \bar{u})^2 = Ns^2$$

with radius $MP = (N)^{1/2}s$ and center at $(\bar{u}, \bar{u}, \cdots, \bar{u})$. Now, the hyperplane (e) will intersect the hypersphere (f) in an $(N-1)$-dimensional hypersphere to correspond to the circle for the case $N = 3$. Consequently, for a given pair of values of $\bar{u}$ and s, the point P will lie on an $(N-1)$-dimensional hypersphere orthogonal to the line OM. The volume of this $(N-1)$-hypersphere is given by

$$A = C_3(\sqrt{N}s)^{N-1}$$

and so

$$dA = C_4 s^{N-2}\, ds.$$

Therefore, the volume $dv = du_1\, du_2 \cdots du_N$ between two concentric spheres of radius $\sqrt{N}s$ and $\sqrt{N}(s + ds)$ is approximately

$$\begin{aligned} dv &= dA\, d(OM) \\ &= Cs^{N-2}\, ds\, d\bar{u}. \end{aligned}$$

Since $du_i = dx_i$ and $d\bar{u} = d\bar{x}$, (17) is established.

7. Distributions of $\bar{x}$, s^2, and s, Taken Singly. It is clear that (18) may be written as follows:

$$(22) \qquad \begin{aligned} dF &= \frac{k}{2} e^{-N\{s^2 + (\bar{x} - \bar{\bar{x}})^2\}/2\sigma^2} (s^2)^{(N-3)/2}\, d(s^2)\, d\bar{x} \\ &= k_1 e^{-N(\bar{x} - \bar{\bar{x}})^2/2\sigma^2}\, d\bar{x} \times k_2 e^{-Ns^2/2\sigma^2} (s^2)^{(N-3)/2}\, d(s^2). \end{aligned}$$

From this factored form it follows that

(a) The law of distribution, $G(\bar{x})$, of sample means from a normal universe is given by

$$(23) \qquad G(\bar{x}) = k_1 e^{-N(\bar{x} - \bar{\bar{x}})^2/2\sigma^2},$$

it being fairly obvious from the form of $G(\bar{x})$ that

$$(24) \qquad k_1 = \left(\frac{2\pi\sigma^2}{N}\right)^{-1/2}.$$

It may also be evaluated by imposing the condition that

$$\int_{-\infty}^{\infty} G(\bar{x})\,d\bar{x} = 1.$$

Evidently, $G(\bar{x})$ is a normal distribution with mean equal to $\tilde{x}$ and standard deviation equal to $\sigma/(N)^{1/2}$, a result already familiar from Chapter VI.

(*b*) The variance, s^2, of a sample is distributed according to

$$(25) \qquad H(s^2) = k_2 e^{-Ns^2/2\sigma^2}\,(s^2)^{(N-3)/2},$$

where (see § 4, Chapter II)

$$(26) \qquad k_2 = \frac{\left(\dfrac{N}{2\sigma^2}\right)^{(N-1)/2}}{\Gamma\left(\dfrac{N-1}{2}\right)}.$$

Thus the distribution of the variance was found by first finding the simultaneous distribution of the variance and mean. Clearly, $H(s^2)$ is a Pearson Type III curve with range limited at one end, $s^2 = 0$, and not at the other, $s^2 = \infty$.

(*c*) The variance, s^2, and the mean, $\bar{x}$, are distributed quite independently, that is,

$$F(\bar{x}, s^2) = G(\bar{x})H(s^2).$$

It has recently been proved by Geary[8] that a necessary and sufficient condition that $\bar{x}$ and s^2 from samples of N values of x be independent in the probability sense is that the x's be normally distributed in the parent universe.

In § 2, the mean of the sampling distribution of s^2 from an arbitrary universe was obtained. It is interesting to verify that result in the present case where we know the distribution function. The mean of the distribution of variances of samples of N from a normal universe is given by

$$E(s^2) = \int_0^{\infty} H(s^2)s^2\,d(s^2),$$

where $H(s^2)$ is defined in (25). So we have

$$E(s^2) = k_2 \int_0^\infty e^{-Ns^2/2\sigma^2}(s^2)^{(N-1)/2}\, d(s^2)$$
$$= k_2 \left(\frac{2\sigma^2}{N}\right)^{(N+1)/2} \Gamma\left(\frac{N+1}{2}\right)$$
$$= \frac{N-1}{N}\sigma^2.$$

The standard deviation of the $H(s^2)$ distribution is, approximately,

$$\sigma_{s^2} = \sqrt{\frac{2}{N}}\,\sigma^2.$$

The distribution of the standard deviations of samples of N from a normal universe is readily found from (25) and (22) to be

$$h(s) = 2k_2 e^{-Ns^2/2\sigma^2} s^{N-2}. \tag{27}$$

So its mean value is given by

$$E(s) = \int_0^\infty h(s)s\, ds$$

which yields the result

$$E(s) = k_2 \left(\frac{2\sigma^2}{N}\right)^{N/2} \Gamma\left(\frac{N}{2}\right).$$

Upon substituting the value of k_2 given in (26), the above expression becomes

$$E(s) = \frac{\left(\frac{2}{N}\right)^{1/2} \Gamma\left(\frac{N}{2}\right)}{\Gamma\left(\frac{N-1}{2}\right)}\sigma. \tag{28}$$

If we denote this coefficient of σ by $b(N)$ we have the relation

$$\sigma = \frac{E(s)}{b(N)}$$

which was promised in § 3. Romanovsky[9] showed that

$$b(N) \to \left(1 - \frac{3}{4N} - \frac{7}{32N^2} - \cdots\right).$$

TABLE 11

N	$1/b(N)$
2	1.772
3	1.382
4	1.253
5	1.189
6	1.151
7	1.126
8	1.108
9	1.094
10	1.084
20	1.040
30	1.026
50	1.015
100	1.008

Table 11 gives values of the reciprocal of $b(N)$ for a few values of N.

Romanovsky also deduced the standard deviation of the $h(s)$ distribution to be

$$\sigma_s = \sigma\left(\frac{1}{2N} - \frac{2}{8N^2} - \frac{3}{16N^3} - \cdots\right)^{1/2}.$$

The approximate value

$$\sigma_s = \left(\frac{1}{2N}\right)^{1/2}\sigma \tag{29}$$

is frequently used in practice and this is the basis for the common statement that *the standard error of a standard deviation is* $1/(2)^{1/2}$ *that of a mean.*

The modal value of s, easily found* by differentiating $h(s)$, is

$$\breve{s} = \sigma\left(\frac{N-2}{N}\right)^{1/2}. \tag{30}$$

If we make the substitution $y = s - \breve{s}$, then the distribution of y is, to a first approximation, the normal curve

$$\text{Const.} \times e^{-Ny^2/2\sigma^2} \tag{31}$$

with standard deviation $\sigma/(2N)^{1/2}$.

8. The $(\bar{x}, s)$-Frequency Surface. We may regard $F(\bar{x}, s)$ as describing a frequency surface if the total volume under the surface represents the expected frequency of the means and standard deviations of all possible samples of size N. In depicting this surface it is convenient to let $\bar{u} = \bar{x} - \tilde{x}$ so that the origin of $\bar{u}$ is at $\bar{x} = \tilde{x}$.

Since

$$\int_0^s \int_{-\infty}^{\infty} F(\bar{x}, s)\, d\bar{x}\, ds = 1,$$

then the volume under the surface over a closed contour in the $\bar{u}s$-plane represents the proportion or percentage of samples whose

* If we make $h(s)$ a maximum for variation in σ we find that
$\sigma = N^{1/2}s/(N-1)^{1/2}$ or $s = \sigma\{(N-1)/N\}^{1/2}$ (cf. Rider[10]).

means and standard deviations fall simultaneously within the ranges defined by the boundary of the given contour. In an illuminating paper[11] by Deming and Birge two such frequency surfaces are represented. These are reproduced in Figure 21, one for a small value of N and the other for a comparatively large value of N.

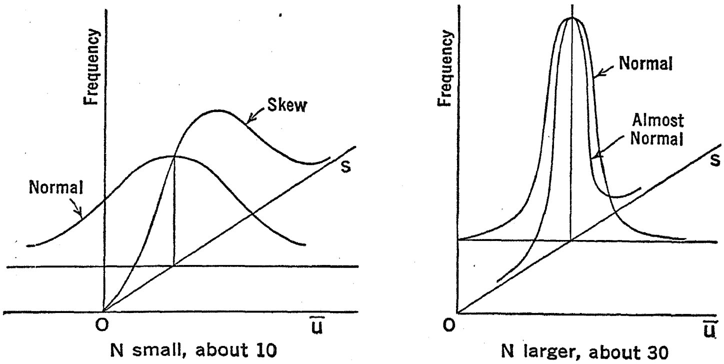


FIG. 21. THE SURFACE $F(\bar{x}, s)$ ILLUSTRATED BY SECTIONS

As the authors point out, the highest point of the surface has the coördinates $\bar{u} = 0$, $s = \sigma\{(N-2)/N\}^{1/2}$. Because of the independence of $\bar{x}$ and s, all plane sections s = constant will be normal curves with standard deviations equal to $\sigma/(N)^{1/2}$. The $\bar{u}$ = constant sections will be skew curves whose equations are given by $h(s)$. They will all have the same mean and mode. As N increases, their mean and mode approach coincidence with the value σ while the curves lose their skewness and become normal with center at $s = \sigma$ and standard deviations equal to $\sigma/(2N)^{1/2}$. As N increases, the surface becomes more and more concentrated about the point $\bar{u} = 0$, $s = \sigma$.

9. Fisher's *t*-Distribution. Substituting (14) into (21) and replacing $N - 1$ by n we obtain

$$F_n(t) = K_n\left(1 + \frac{t^2}{n}\right)^{-(n+1)/2}, \tag{32}$$

where $1/K_n = n^{1/2}\mathrm{B}(n/2, 1/2)$, B being the Beta function.

Inasmuch as (32) is independent of σ, it can be used in situations in which the value of σ is unknown. The quantity t involves no hypothetical quantities, being completely expressible in terms of the variates.

In 1925, "Student" published in *Metron*[12] an extensive table of the probability integral $\int_{-\infty}^{t} F_n(t)\,dt$. More recently, Fisher[13] has given a short table of the probability P of occurrence of deviations outside $\pm t$, for values of t and n commonly met in applications of small sample theory. Let

$$P_n(t) = 2\int_0^t F_n(t)\,dt.$$

Then the probability P tabulated by Fisher is

$$P = 1 - P_n(t). \tag{33}$$

Fisher's table gives successive columns showing for each value of n, from $n = 1$ to $n = 30$, the values of t for which P takes the values given at the head of the columns. A general idea of the table may be obtained from the portion which we have reproduced* in Table 12.

TABLE 12. VALUES OF t FROM TABLE IV OF FISHER'S TEXT

n \ P	.9	.7	.5	.1	.05	.01
3	.137	.424	.765	2.353	3.182	5.841
4	.134	.414	.741	2.132	2.776	4.604
5	.132	.408	.727	2.015	2.571	4.032
6	.131	.404	.718	1.943	2.447	3.707
8	.130	.399	.711	1.860	2.306	3.355
10	.129	.397	.706	1.812	2.228	3.169
15	.128	.393	.691	1.753	2.131	2.947
20	.127	.391	.687	1.725	2.086	2.845
30	.127	.389	.683	1.697	2.042	2.750
∞	.1257	.3853	.6745	1.6449	1.9600	2.5758

The number n, with which to enter the table, is determined by the number of degrees of freedom involved in the available estimate (§ 3) of σ^2. In testing null hypotheses the rule given in § 12 of Chapter VI may be used, where, of course, P_δ is to be replaced now by P.

The distribution of t (as well as that of z) approaches the normal type as $n \to \infty$. This may be established as follows. Using Stirling's approximation on the coefficient K_n in (32), we obtain, after some

* With Fisher's permission and that of his publishers, Oliver and Boyd.

algebraic simplification, the following expression:

$$K_n = e^{-1/2}(n\pi)^{-1/2}\left(\frac{n-1}{n-2}\right)^{(n-2)/2}\left(\frac{n-1}{n-2}\right)^{1/2}\left(\frac{n-1}{2}\right)^{1/2}.$$

From this it is easy to show that

$$\lim_{n\to\infty} K_n = (2\pi)^{-1/2}.$$

The rest of the t function may be written as

$$\left(1+\frac{t^2}{n}\right)^{-1/2}\left(1+\frac{t^2}{n}\right)^{-n/2}$$

which, when $n = \infty$, becomes $e^{-t^2/2}$. Therefore,

$$\lim_{n\to\infty} F_n(t) = (2\pi)^{-1/2}e^{-t^2/2}.$$

The entries in the last line of Fisher's table, corresponding to $n = \infty$, are the deviations from the mean of a normal curve with unit standard deviation.

According to " Student," the distribution of z tends to approach a normal curve with a standard deviation of $(N-3)^{-1/2}$ for large values of N. Deming and Birge (*loc. cit.*) have suggested that the distribution tends to approach normality with $(N - 3/2)^{-1/2}$ as standard deviation. Anyhow, for large values of N, $(N-3)^{1/2}z$ would be approximately normally distributed about zero with unit standard deviation. Since

$$(N-3)^{1/2}z = \left\{\left(\frac{N-3}{N-1}\right)\right\}^{1/2} t \quad \text{and} \quad t = \frac{(\bar{x}-\tilde{x})(N-1)^{1/2}}{s},$$

it is frequently satisfactory in applications to refer

$$t = \frac{(\bar{x}-\tilde{x})(N-3)^{1/2}}{s} \tag{34}$$

to a normal probability scale when $N > 30$.

For large values of N, (34) represents so small a refinement over (22) of Chapter VI that the additional computation seems unwarranted. So when N considerably exceeds 30 the older procedure of replacing σ by s and treating $t = (\bar{x}-\tilde{x})(N)^{1/2}/s$ as though it were normally distributed with unit standard deviation is not appreciably erroneous.

10. Difference Between Two Means. Fisher[7] demonstrated that (32) has a much wider range of application than the problem for which it was designed. He showed that the t-distribution is applicable whenever we are dealing with a normally distributed variate whose standard deviation is not known exactly but is independently estimated from observations amounting to n degrees of freedom. The scheme by which the " Student " idea is made available to other problems consists in constructing a variable t in the nature of a fraction whose numerator is any statistic normally distributed and whose denominator is the square root of an independently distributed and unbiased estimate of the variance of the numerator involving n degrees of freedom. Thus the t-distribution has been found useful in such problems as testing the significance of the difference between two means and testing hypotheses regarding regression coefficients.

Let $\bar{x}_1$, $\bar{x}_2$ be the means and s_1, s_2 the standard deviations of two independent samples of N_1 and N_2 variates, respectively, from a normal universe with mean $\bar{x}$ and variance σ^2. According to (10) of Chapter VI the variance of the difference between the two means is $\sigma^2 (N_1 + N_2)/N_1N_2$. Then it can be proved[14] that the variable

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sigma}\left\{\frac{N_1N_2}{N_1 + N_2}\right\}^{1/2} \tag{35}$$

is normally distributed with unit standard deviation. However, in most practical problems σ is unavailable and must be estimated from the samples. Using the unbiased estimate defined in (5), the above formula becomes

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\hat{\sigma}}\left\{\frac{N_1N_2}{N_1 + N_2}\right\}^{1/2}. \tag{36}$$

Fisher showed that (36) is distributed in accord with (32) for $n = N_1 + N_2 - 2$, and we can find from Fisher's table of P the probability of a greater difference between the means than that observed.

As N_1 and N_2 become large, $(N_1 + N_2)/(N_1 + N_2 - 2)$ tends toward unity and (36) tends toward the value

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\left\{\dfrac{s_1^2}{N_2} + \dfrac{s_2^2}{N_1}\right\}^{1/2}}. \tag{37}$$

Since (36) is asymptotically normally distributed, the older procedure of referring (37) to a normal probability scale in testing a null hypothesis that two samples are from the same universe would not be invalid to any appreciable extent for large values of N_1 and N_2. The present writer[15] has recently called attention to an erroneous formula which is commonly used in place of (37).

If one of the samples, say N_2, is so much larger than the other that it tends toward the universe, then $\bar{x}_2$ tends toward $\tilde{x}$ and s_2 tends toward σ. So, under these conditions, (37) tends toward

$$t = \frac{(\bar{x}_1 - \tilde{x})\sqrt{N_1}}{\sigma}$$

which, if the subscripts are dropped, is the formula used in testing a null hypothesis that a given sample comes from a normal universe with a proposed mean. When $N_1 = N_2 = N$, (36) reduces to

$$t = (\bar{x}_1 - \bar{x}_2)\left\{\frac{N-1}{s_1^2 + s_2^2}\right\}^{1/2}. \tag{38}$$

Inasmuch as we do not ordinarily know whether a sample is drawn from a normal universe or some other type of universe, a question quite naturally arises as to whether the procedure inaugurated by " Student " and extended by Fisher is applicable to small samples from non-normal universes. The question may be considered partially answered by Bartlett[16] and others who have shown that it gives a good approximation for considerable departures from normality in the sampled universe. However, a word of caution seems to be in order lest the new procedure be oversold in the applications by completely neglecting the underlying assumptions of normality in the universe and randomness of the samples.

The following examples, cited by Rietz,[17] illustrate the " Student " theory.

Example 1. Suppose a random sample of $N = 5$ is obtained from a hypothetical normal universe whose mean is $\tilde{x} = 2$. It is found that $\bar{x} = 3$ and $s^2 = \frac{4}{5}$ for the sample. What is the probability that one would obtain a sample of five for which $\bar{x}$ would differ numerically from $\tilde{x}$ by as much as unity?

Solution. From (12), $t = \sqrt{5} = 2.236$. Entering Fisher's table for $n = 4$, we find the probability P between .1 and .05. Reference to the more extensive table in *Metron*[12] gives $P = .0892$ for the probability of a discrepancy as large as the one observed. It is interesting to compare this result with what would be obtained by reference to a normal probability scale. We find $P = .0254$ for a deviation outside $t = \pm 2.236$. In terms of the odds that a mean, $\bar{x}$, will

deviate numerically more than 1 from theory, the contrast is more striking. Thus, under the "Student" theory we should say that the odds are 10,000 to 892 or roughly 11 to 1 against a deviation as large as or larger than the one observed. Under the normal theory the odds are 10,000 to 254, or about 40 to 1 against such a deviation.

Example 2. The following data represents the yields in bushels of Indian corn on ten subdivisions of equal areas of two agricultural plots in which Plot 1 was a control plot treated the same as Plot 2, except for the amount of phosphorus applied as a fertilizer.

Plot 1	*Plot* 2
6.2	5.6
5.7	5.9
6.5	5.6
6.0	5.7
6.3	5.8
5.8	5.7
5.7	6.0
6.0	5.5
6.0	5.7
5.8	5.5
10 |60.0	10 |57.0
$\bar{x}_1 = 6.0$	$\bar{x}_2 = 5.7$

Is there a significant difference between the yields on the two plots, using the difference between their means as a criterion of judgment?

Solution.

$$s_1^2 = \frac{.64}{10} = .064$$

$$s_2^2 = \frac{.24}{10} = .024.$$

Substitution in (38) gives

$$t = (6.0 - 5.7)\left\{\frac{9}{.088}\right\}^{1/2}$$

$$= (.3)(10.113) = 3.034.$$

Entering "Student's" tables in *Metron* (*loc. cit.*) at $n = 18$, we find $P = .0072$ for the probability that t will fall outside the range -3.034 and $+3.034$. Hence a null hypothesis that the samples are from the same universe would be refuted by the test for both the .05 and .01 levels of significance. In other words, our conclusion is that, on the levels of significance adopted, there is a significant difference between the yields on the plots.

11. Fisher's z-Distribution. Suppose u^2 and v^2 are two independent and unbiased estimates of the variance σ^2 of a variable x which is normally distributed. If these estimates are based upon samples of N_1 and N_2, respectively, or upon n_1 and n_2 degrees of

freedom, then we have

$$u^2 = \frac{1}{N_1 - 1}\sum_1^{N_1}(x_{1i} - \bar{x}_1)^2 = \frac{1}{n_1}\sum_1^{n_1+1}(x_{1i} - \bar{x}_1)^2$$

$$v^2 = \frac{1}{N_2 - 1}\sum_1^{N_2}(x_{2i} - \bar{x}_2)^2 = \frac{1}{n_2}\sum_1^{n_2+1}(x_{2i} - \bar{x}_2)^2$$

in which $\bar{x}_1$ and $\bar{x}_2$ are the means of the two samples. In previous notation u^2 and v^2 would be denoted by $\hat{\sigma}_1{}^2$ and $\hat{\sigma}_2{}^2$ but these symbols are too unwieldy in the present discussion.

In constructing a test of significance for the difference between two sample variances it might seem logical to form the difference $w = u^2 - v^2$ and seek the distribution function of w. However, such a procedure is impractical because of the mathematical difficulty involved in determining this function. Fisher circumvented this difficulty by building a statistic, z, defined by

$$z = \tfrac{1}{2}(\log_e u^2 - \log_e v^2) = \log_e \frac{u}{v} \tag{39}$$

whose distribution function, $G(z)$, he obtained and which proved to have extremely wide application. To derive $G(z)$ we make use of the distribution of $H(s^2)$ given in (25), replacing $N - 1$ by n and s^2 by $(n/\overline{n + 1})\, u^2$ (see § 3). After this modification, (25) becomes

$$\frac{\left\{\dfrac{n}{2\sigma^2}\right\}^{n/2}}{\Gamma\left(\dfrac{n}{2}\right)} (u^2)^{(n-2)/2} e^{-nu^2/2\sigma^2}\, d(u^2). \tag{40}$$

Since u^2 and v^2 are independent their joint distribution is

$$K\,(u^2)^{(n_1-2)/2}(v^2)^{(n_2-2)/2}e^{-(n_1u^2+n_2v^2)/2\sigma^2}\, d(u^2)\, d(v^2) \tag{41}$$

where

$$K = \frac{(n_1)^{n_1/2}(n_2)^{n_2/2}}{2^{(n_1+n_2)/2}\sigma^{(n_1+n_2)}\Gamma\left(\dfrac{n_1}{2}\right)\Gamma\left(\dfrac{n_2}{2}\right)}.$$

From (39) we have

$$u^2 = v^2 e^{2z} \tag{42}$$

and for a fixed value of v^2,

$$d(u^2) = 2v^2e^{2z}\, dz. \tag{43}$$

Using (42) and (43) in (41) we obtain

$$(44) \qquad 2Ke^{n_1z}e^{-(n_1e^{2z}+n_2)v^2/2\sigma^2}(v^2)^{(n_1+n_2-2)}\,d(v^2)\,dz$$

for the joint distribution of v^2 and z. Integrating (44) with respect to v^2 between the limits 0 and ∞ and making use of the Gamma function we obtain the distribution of z,

$$(45) \qquad G(z) = \frac{2n_1^{n_1/2}n_2^{n_2/2}}{B\left(\frac{n_1}{2}, \frac{n_2}{2}\right)} \frac{e^{n_1z}\,dz}{(n_1e^{2z} + n_2)^{(n_1+n_2)/2}}.$$

The function $G(z)$ has the important property that it depends solely upon n_1 and n_2, not at all upon the variance of the sampled universe. Fisher's z should not be confused with the z-distribution of "Student."

The distribution function of z is extremely general, including as special cases, the χ^2-distribution, the t-distribution of "Student" and Fisher, and the normal distribution. Rider[18] has made easily available the transformations and substitutions by which these special cases can be obtained from (45).

The positive part of the curve for $z = \log_e(u/v)$ is the same as the negative part for $z = \log_e(v/u)$. Since it is optional which estimate is considered as u^2 it is necessary, in tabulating the probability integral of $G(z)$, to consider only positive values of z by making u^2 the larger variance estimate (based on n_1 degrees of freedom).

Let $Q = \int_{-\infty}^{z_0} G(z)\,dz$ and let $P = 1 - Q$. Thus P is the probability that $z > z_0$. In his book, Fisher[19] has given values of z_0 corresponding to the probabilities $P = .05$ and $.01$ for various combinations of n_1 and n_2. These values, z_0, are called the "5% and 1% points" and are used as critical values in judging significance. It should be noticed that Fisher's "points" are based on the area of the *whole* curve and therefore they should not be confused with 5% and 1% "levels of significance" previously used. In the latter sense, Fisher's "points" would be 10% and 2% "levels of significance." In other words, a 5% point means a value of z such that one "tail" under the curve is .05, whereas a 5% level of significance meant a value of t such that the sum of both "tails" (outside $\pm t$) is .05. It is hoped that tables of 5% and 1% levels of significance for z will sometime be available.

12. Significance of Difference Between Variances. The usual hypothesis tested by the z-test is that u^2 and v^2 are estimates of one and the same population variance and therefore that $z = 0$. The significance of the divergence of the observed value of z from zero is the crux of the test. Small values of z mean a tenable hypothesis whereas values of z larger than z_0 refute the hypothesis. If for $P = .05$ (or $.01$) the observed value of z, as computed from the samples in accordance with (39), is larger than z_0, the hypothesis is to be rejected and the conclusion is that the samples come from universes with different variances.

Logically, the z-test should be applied before testing the difference between two means since the latter test depends on the equality of the population variances.

To avoid the troublesome logarithmic computation involved in (39) Snedecor[20] has published tables which transform Fisher's 5% and 1% points into the ratio u^2/v^2, where $e^{2z} = u^2/v^2$. Snedecor calls this ratio F in honor of Fisher.* Therefore,

$$F = \frac{u^2}{v^2}$$

where u^2 is to be chosen the larger of the two given variance estimates. This table is reproduced in the Appendix. (See Table II.)

Example 3. In Example 2 suppose we wish to test the assumption, which was made there, that the two samples come from universes with equal variance. We have

$$u^2 = \frac{n_1 + 1}{n_1} s_1{}^2 = \frac{.64}{9} = .0711$$

$$v^2 = \frac{n_2 + 1}{n_2} s_2{}^2 = \frac{.24}{9} = .0267$$

$$F = \frac{.0711}{.0267} = 2.663$$

$$\begin{aligned} z &= .5 \log_e F \\ &= 1.1513 \log_{10} F = .49. \end{aligned}$$

Entering Fisher's table (*loc. cit.*) for $n_1 = n_2 = 9$ we find $z_0 = .58$ for $P = .05$ and $z_0 = .84$ for $P = .01$. This means that, if the true value of z were zero, random sampling fluctuations would be expected to give a value of z as great as .84, or greater, once in 100 trials, and a value of z as great as .58, or greater, five

* In their new *Statistical Tables* Fisher and Yates call it the *variance ratio.* These tables are published by Oliver and Boyd, London.

times in 100 trials. The observed value of z is .49 and so this value might be accounted for by chance, at either the .05 or .01 points of significance. Using Snedecor's table we find $F = 3.18$ for $P = .05$ and $F = 5.35$ for $P = .01$. Since the observed value of F is only 2.663, we conclude that we were justified in proceeding with the t-test.

When the samples are large there are two procedures available.

I. $G(z)$ is skew when $n_1 \neq n_2$ but when $n_1 = n_2$ it is symmetrical. When n_1 and n_2 are large and also for moderate values when they are equal or nearly equal one can verify (by taking logarithms) that z is approximately normally distributed about zero with mean zero and variance $\frac{1}{2}(1/n_1 + 1/n_2)$. Therefore,

$$t = \frac{z}{\left\{\frac{1}{2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\right\}^{1/2}} \tag{46}$$

may be referred to a normal probability scale.

II. Let $w = s_1 - s_2$. From (10) of Chapter VI and (29) of this chapter,

$$\sigma_w = \sigma\left(\frac{1}{2N_1} + \frac{1}{2N_2}\right)^{1/2}.$$

Then

$$t = \frac{s_1 - s_2}{\sigma\left(\frac{1}{2N_1} + \frac{1}{2N_2}\right)^{1/2}} \tag{47}$$

is normally distributed about zero with unit standard deviation. An estimate of the supposed common variance σ^2 is given in (5). Using the square root of this estimate in place of σ in (47) and assuming that N_1 and N_2 are large enough to regard, without appreciable error, the ratio $(N_1 + N_2)/(N_1 + N_2 - 2)$ as unity we obtain

$$t = \frac{s_1 - s_2}{\left\{\frac{s_1^2}{2N_2} + \frac{s_2^2}{2N_1}\right\}^{1/2}}. \tag{47a}$$

This value may then be referred to a normal probability scale.

An interesting derivation, using characteristic functions, of a method for testing the significance of the difference between two sample variances has recently been given by A. T. Craig.[21]

13. Analysis of Variance. The test of significance between two independent sample variances (with their appropriate degrees of freedom) is a special case of a general technique, developed by Fisher, for segregating the variance into portions traceable to specific sources. In general, the kind of procedure one attempts to follow in such an analysis can be illustrated by the following scheme.

Let us imagine a individuals $I_1, I_2, \cdots, I_a$, each subjected to b treatments $T_1, T_2, \cdots, T_b$. For example, the I's may be agricultural plots containing different varieties of some plant and the T's may be applications of various kinds or amounts of fertilizers. Or the I's might conceivably be various diabetic patients and the T's varietal insulin treatments. The effects of the T's on the I's yield a set of observations, to be denoted by x_{jk}, which vary from one value of I to another for a fixed T and from one value of T to another for a fixed I. Suppose, then, that $N = ab$ independently observed values of a normally distributed variable are classified into a rows and b columns in accordance with some relevant scheme as depicted in Table 13.

TABLE 13. MATRIX OF $N = ab$ INDEPENDENT VALUES FROM A NORMAL UNIVERSE

	T_1	T_2	$\cdots$	T_b
I_1	x_{11},	x_{12},	$\cdots$	x_{1b}
I_2	x_{21},	x_{22},	$\cdots$	x_{2b}
..	. . . .	. . . .	. . . .	.
I_a	x_{a1},	x_{a2},	$\cdots$	x_{ab}

The values in each row will vary about the mean of that row and the values in each column will vary about the mean of that column. Let $\bar{x}_{j\cdot}$ denote the mean of the jth row,

$$(48) \qquad b\bar{x}_{j\cdot} = \sum_{k=1}^{b} x_{jk}, \qquad j = 1, 2, \cdots, a,$$

and let $\bar{x}_{\cdot k}$ denote the mean of the kth column,

$$(49) \qquad a\bar{x}_{\cdot k} = \sum_{j=1}^{a} x_{jk}, \qquad k = 1, 2, \cdots, b.$$

(The dot indicates that summation has been effected on the index

which it replaces.) Let the mean of the entire set be $\bar{x}$ where

$$ab\bar{x} = \sum_1^a \sum_1^b x_{jk}. \tag{50}$$

Let the variance in the entire set due to all causes be Q/ab where

$$Q = \sum_1^a \sum_1^b (x_{jk} - \bar{x})^2 = \sum_1^{ab} (x_i - \bar{x})^2. \tag{51}$$

Now Q can be resolved into three quadratic forms as follows:

$$Q = q_1 + q_2 + q_3 \tag{52}$$

where

$$q_1 = b\sum_1^a (\bar{x}_{j\cdot} - \bar{x})^2$$

$$q_2 = a\sum_1^b (\bar{x}_{\cdot k} - \bar{x})^2$$

$$q_3 = \sum_1^a \sum_1^b (x_{jk} - \bar{x}_{j\cdot} - \bar{x}_{\cdot k} + \bar{x})^2.$$

That (52) is an identity in the $N = ab$ values of x can be readily seen as follows:

$$\begin{aligned}\sum_1^a \sum_1^b (x_{jk} - \bar{x})^2 &= \sum_1^a \sum_1^b \{(x_{jk} - \bar{x}_{j\cdot} - \bar{x}_{\cdot k} + \bar{x}) + (\bar{x}_{j\cdot} - \bar{x}) + (\bar{x}_{\cdot k} - \bar{x})\}^2 \\ &= \sum_1^a \sum_1^b (x_{jk} - \bar{x}_{j\cdot} - \bar{x}_{\cdot k} + \bar{x})^2 + \sum_1^a \sum_1^b (\bar{x}_{j\cdot} - \bar{x})^2 \\ &\qquad + \sum_1^a \sum_1^b (x_{\cdot k} - \bar{x})^2.\end{aligned}$$

To show that the cross-product terms vanish consider the term

$$\sum_1^a \sum_1^b (x_{jk} - \bar{x}_{j\cdot} - \bar{x}_{\cdot k} + \bar{x})(\bar{x}_{j\cdot} - \bar{x}).$$

This becomes

$$\begin{aligned}&\sum_{j=1}^a (\bar{x}_{j\cdot} - \bar{x}) \sum_{k=1}^b (x_{jk} - \bar{x}_{j\cdot} - \bar{x}_{\cdot k} + \bar{x}) \\ &\qquad = \sum_{j=1}^a (\bar{x}_{j\cdot} - \bar{x})(b\bar{x}_{j\cdot} - b\bar{x}_{j\cdot} - b\bar{x} + b\bar{x}) = 0.\end{aligned}$$

A similar demonstration can be made for the other cross-product

terms. This is left as an exercise for the student. Since

$$\sum_1^a\sum_1^b(\bar{x}_{j\cdot} - \bar{x})^2 = b\sum_1^a(\bar{x}_{j\cdot} - \bar{x})^2$$

$$\sum_1^a\sum_1^b(\bar{x}_{\cdot k} - \bar{x})^2 = a\sum_1^b(x_{\cdot k} - \bar{x})^2$$

(52) is established.

The variability *between rows* is measured by q_1 and *between columns* by q_2. The residual variability, freed from the influence of either rows or columns, is measured by q_3 and is called *interaction* (sometimes also *discrepance*). It may be regarded as the "experimental error" inherent in the experiment and over which no control is attempted. As will be shown later, it is used as a standard against which the variability measured by either q_1 or q_2 may be tested for significance, when the appropriate number of degrees of freedom are taken into account.

From (51) the number of degrees of freedom in Q is seen to be $N - 1$. Since there are a values of $\bar{x}_{j\cdot}$ the number of degrees of freedom in q_1 is $(a - 1)$. Similarly, the number in q_2 is $(b - 1)$. This leaves $(N - 1) - \{(a - 1) + (b - 1)\} = (a - 1)(b - 1)$ for q_3, a result which may also be deduced from the expression for q_3. Another form of argument is as follows. The ab means of rows and columns form an $(a \times b)$-fold table of $(a - 1)(b - 1)$ degrees of freedom since the marginal means are fixed in terms of the x_{jk} values. Anyhow, the number of degrees of freedom in interaction is the product of the numbers in the interacting forces. Accordingly, an unbiased estimate of σ^2 from the rows is $q_1/(a - 1)$, from the columns is $q_2/(b - 1)$, and from interaction is $q_3/(a - 1)(b - 1)$. It is clear, therefore, that the z-distribution can be employed to test the significance of the variability attributable to these sources if the independence of the above-mentioned estimates is assured. A. T. Craig[21] has settled this point by establishing the independence of the q's.

The quantities required in an analysis of variance are summarized in Table 14. They can be readily computed except possibly q_3. So long as the arithmetic involved in computing the other quantities is carefully checked it is sufficient to evaluate q_3 from relation (52). In other words, the sum of squares due to *interaction* may be found by subtracting $(q_1 + q_2)$ from the *total* sum of squares.

TABLE 14

Variance due to	*D. of F.*	*Sum of Squares*	*Unbiased Estimates*
Rows	$a - 1$	$q_1 = b\sum_1^a (\bar{x}_{j\cdot} - \bar{x})^2$	$q_1/(a - 1)$
Columns	$b - 1$	$q_2 = a\sum_1^b (\bar{x}_{\cdot k} - \bar{x})^2$	$q_2/(b - 1)$
Interaction	$(a - 1)(b - 1)$	$q_3 = Q - q_1 - q_2$	$q_3/(a - 1)(b - 1)$
Total	$ab - 1$	$Q = \sum_1^a \sum_1^b (x_{jk} - \bar{x})^2$	

Under the null hypothesis that there is no significant variation from row to row, the quantity

$$z = \tfrac{1}{2} \log_e \frac{(b - 1)q_1}{q_3} \tag{53}$$

will be distributed in accord with (45) and the hypothesis can be tested from critical values of z or, more conveniently, perhaps, from Snedecor's table by computing

$$F = \frac{(b - 1)q_1}{q_3} \tag{54}$$

and entering the table at (n_1, n_2) where $n_1 = b - 1$, and $n_2 = (a - 1)(b - 1)$. If the computed value falls above the critical value adopted, the null hypothesis is rejected for that value. Similarly, to test the null hypothesis that there are no significant effects from column to column we compute

$$F = \frac{(a - 1)q_2}{q_3} \tag{55}$$

and compare it with one of the tabular entries for $n_1 = a - 1$, $n_2 = (a - 1)(b - 1)$.

Example 4. On a feeding experiment a farmer has four types of hogs denoted by I, II, III, IV. These types are each divided into three groups which are fed varietal rations A, B, and C. The following results are obtained, the numbers in the table being the gains in weight in pounds in the various groups.

	I	II	III	IV	Totals
A	7.0	16.0	10.5	13.5	47.0
B	14.0	15.5	15.0	21.0	65.5
C	8.5	16.5	9.5	13.5	48.0
Totals	29.5	48.0	35.0	48.0	160.5

The computations yield the following results:

Sum of Squares		*D. of F.*	*Unbiased Estimates*
Rations	54.1250	2	27.06
Types	87.7292	3	29.24
Interaction	28.2083	6	4.70

To test the significance of the variation in rations we refer $F = 27.06/4.70 = 5.76$ to Snedecor's table where, corresponding to (2, 6) degrees of freedom, we find 5.79 for the 5% point and 10.92 for the 1% point. Similarly, to test the significance of the variation between types we compute $F = 29.24/4.70 = 6.2$. The entries in the table for (3, 6) degrees of freedom are 4.76 for the 5% point and 9.78 for the 1% point. Our conclusion is that there is a significant difference between breeds (somewhat doubtful) and between varieties of rations at the 5% point, but that neither is significant at the 1% point.

14. Testing Variation in Sub-sets of Means. In a previous chapter a method was given for testing the significance of a difference between two means. We shall now show that the analysis of variance technique lends itself to testing the significance of differences between any number of group means.

Consider normal universes with means $\tilde{y}_x$, $(x = 1, 2, \cdots, b)$, and variance σ^2. Let samples of N_x be drawn one from each of these universes and let $\bar{y}_x$ and s_x^2 be the mean and variance of the sample of N_x. Thus we have b classes or arrays (as in a correlation table). The notation for the samples is summarized in Table 15.

TABLE 15

Classes	1	2	$\cdots$	x	$\cdots$	b
Means	$\bar{y}_1,$	$\bar{y}_2,$	$\cdots$	$\bar{y}_x,$	$\cdots$	$\bar{y}_b$
Standard Deviations	$s_1,$	$s_2,$	$\cdots$	$s_x,$	$\cdots$	s_b
Frequencies	$N_1,$	$N_2,$	$\cdots$	$N_x,$	$\cdots$	N_b

Our problem is to test, from the samples, the hypothesis that $\tilde{y}_1 = \tilde{y}_2 = \cdots = \tilde{y}_b$.

It can be shown (Cf. Part I; Ex. 3, p. 208) that the sum of the squares of deviations of the $N = \sum_1^b N_x$ variates y_x from the mean $\bar{y}$ of the entire set may be broken up into two parts such that

$$V = v_1 + v_2$$

where

$$V = \sum_1^N (y_x - \bar{y})^2$$

$$v_1 = \sum_1^b N_x(\bar{y}_x - \bar{y})^2$$

$$v_2 = \sum_1^b N_x s_x^2.$$

It is conventional to call v_1 the variation *between classes* and v_2 the variation *within classes.*

An unbiased estimate of $\tilde{y}_x$ is $\bar{y}$ where $N\bar{y} = \sum_1^b N_x \bar{y}_x$. Hence there are $b - 1$ degrees of freedom in v_1. An unbiased estimate of σ^2 from the values of $\bar{y}_x$ is $v_1/(b - 1)$, and from the values of s_x^2 is $v_2/(N - b)$ since the variates in the computation of s_x^2 are subject to the linear restriction $\sum_1^{N_x} y_x = N_x \bar{y}_x$ and there are b values of x. Therefore, under the null hypothesis that $\tilde{y}_1 = \tilde{y}_2 \cdots = \tilde{y}_b$, the quantity

$$z = \tfrac{1}{2} \log_e \frac{(N - b)v_1}{(b - 1)v_2} \tag{56}$$

is distributed in accord with (45) and the hypothesis can be tested by computing

$$F = \frac{(N - b)v_1}{(b - 1)v_2} \tag{57}$$

and comparing it with the entries in Snedecor's table for (n_1, n_2) where $n_1 = b - 1$, $n_2 = N - b$. The quantities required in the computations are summarized in Table 16.

TABLE 16

Variance due to	*D. of F.*	*Sum of Squares*	*Unbiased Estimates*
Between Classes	$b - 1$	v_1	$v/(b - 1)$
Within Classes	$N - b$	v_2	$v/(N - b)$
Total	$N - 1$	V	$\dfrac{(N - b)v_1}{(b - 1)v_2} = F$

The variation within classes is independent of the principle of classification. Therefore, excessive variation between classes (variation of the $\bar{y}_x$'s) as compared with variation within classes (variation of sample values about their respective means) will cause F to fall above the critical value adopted, and the null hypothesis is contradicted or refuted for that value.

Examples from agricultural and certain branches of biological science will be found in the textbooks by Fisher and by Snedecor, and from the field of economics in Mills' text (revised edition).

15. Testing Linear Regression. Consider a correlation table with b arrays in the x direction. Let $f(x)$ represent the frequency and $\bar{y}_x$ the mean in the array at x. Let $(\bar{x}, \bar{y})$ be the mean of the table and m_1 and m_2 the linear regression coefficients as defined in Part I. Suppose the $N = \sum_x \sum_y f(x, y)$ entries in the table constitute a sample from a normal bivariate universe and we wish to test the hypothesis, H, that the regression of y on x is linear. It is shown in Part I that $Y_x - \bar{y} = m_1(x - \bar{x})$ is the equation of the line which fits the means of the arrays best, in a least-squares sense, and so Y_x is the estimated mean of the array at x. (A slightly different notation was used in Part I.)

The variation B *between* arrays can be resolved into two components B_1 and B_2 such that

$$B = B_1 + B_2 \tag{58}$$

where

$$B = \sum_1^b f(x)(\bar{y}_x - \bar{y})^2$$

$$B_1 = \sum_1^b f(x)(\bar{y}_x - Y_x)^2$$

$$B_2 = \sum_1^b f(x)(Y_x - \bar{y})^2.$$

To establish (58) we may write B in the form

$$\sum_1^b f(x)\{(\bar{y}_x - Y_x) + (Y_x - \bar{y})\}^2$$

which upon expansion equals $B_1 + B_2$ because, as the student may verify, the cross-product term vanishes.

It is shown in Part I that $B = N\eta_{yx}^2\sigma_y^2$ (Cf. (39), p. 200) and $B_2 = Nr^2\sigma_y^2$ (Cf. (16), p. 172). Since B_2 is the part of B which is accounted for by H it follows from (58) that $B_1 = N\sigma_y^2\ (\eta_{yx}^2 - r^2)$ is the part of B not accounted for by H. We are interested in the question, Is B_1 excessive compared with the random sampling fluctuations to be expected under the null hypothesis? To answer this question consider the variation W *within* arrays where

$$W = \sum_x f(x)(y - \bar{y}_x)^2.$$

In Part I this was designated by NS'^2_y which in turn is equal to $N\sigma_y^2(1 - \eta_{yx}^2)$. This variation within classes is due to a host of random forces which are not dependent on the value of x defining the arrays. Therefore, W provides a basis for testing whether B_1 is small enough to be accepted as the resultant of random forces under H or whether it is so large as to contradict H. Before we can use the z-test, however, the degrees of freedom must be reckoned. In B there are $b - 1$ degrees of freedom because the b values of $\bar{y}_x$ are subject to the linear restriction $\sum_{x=1}^b \bar{y}_x f(x) = N\bar{y}$. The number in B_2 may be determined by making use of the regression equation and writing B_2 in the form

$$\sum_{x=1}^b f(x)(Y_x - \bar{y}) = m_1^2 \sum_{x=1}^b f(x)(x - \bar{x})^2.$$

Since $\sum_{x=1}^{b} f(x)(x - \bar{x})^2$ is independent of the regression, the variation in B_2 must be due to the single statistic m_1 and therefore involve one degree of freedom. Hence, from (58), there are $b - 2$ degrees of freedom in B_1. Since there are b arrays there are $N - b$ degrees of freedom in W. Consequently,

$$z = \tfrac{1}{2} \log_e \frac{\eta^2 - r^2}{1 - \eta^2} \frac{N - b}{b - 2}$$

is distributed in accord with (45) if H is true. The computed value of

$$F = \frac{\eta^2 - r^2}{1 - \eta^2} \frac{N - b}{b - 2}$$

may, therefore, be compared with one of the entries in Snedecor's table for $n_1 = b - 2$, $n_2 = N - b$.

This is the test which was promised in Part I to replace the Blakeman criterion which Fisher proved was unsound. The student may construct a similar argument for testing an hypothesis of linear regression of x on y.

16. Tests of Significance of r. Let the variables x, y be simultaneously distributed in accord with some one or other of the distribution functions

$$f(x, y) = Ke^{-P}, \qquad -\infty \leq x \leq \infty, -\infty \leq y \leq \infty,$$

where

$$\frac{1}{K} = 2\pi\sigma_x\sigma_y(1 - \rho^2)^{1/2}$$

$$P = \frac{1}{2(1 - \rho^2)} \left\{ \frac{(x - \tilde{x})^2}{\sigma_x^2} - \frac{2\rho(x - \tilde{x})(y - \tilde{y})}{\sigma_x\sigma_y} + \frac{(y - \tilde{y})^2}{\sigma_y^2} \right\},$$

and $\tilde{x}$, $\tilde{y}$, σ_x, σ_y, and ρ are undetermined. In other words, suppose that the universe is *some* normal bivariate distribution. The question of the reliability of a value of r computed from a sample of N pairs of (x, y) from such a universe may conveniently be discussed under two cases.

Case I. When $\rho = 0$. In testing the significance of an observed value of r we are testing the hypothesis that $\rho = 0$. Under this hypothesis the sampling distribution of r is known to be

$$f(r) = k(1 - r^2)^{(N-4)/2}, \qquad -1 \leq r \leq 1,$$

where $1/k = B(\frac{1}{2}, \overline{N - 1}/2)$. The curves represented by this function are symmetrical about $r = 0$ with

$$\sigma_r = (N - 1)^{-1/2}$$

$$\alpha_4 = 3 - \frac{6}{N - 1}.$$

As N becomes large the function is practically normal and consequently

$$t = r(N - 1)^{1/2} \tag{59}$$

tends to be normally distributed with mean zero and unit standard deviation. Therefore, to test the significance of a value of r computed from a large sample it would not be invalid, to any appreciable extent, to refer (59) to a normal probability scale.

When N is small the problem may be resolved into an analysis of variance. In a correlation table, the total variation in the y direction may be broken up into two parts, (1) the part $Nr^2\sigma_y{}^2$ which may be accounted for by an hypothesis of linear regression and (2) the residual part $NS_y{}^2 = N\sigma_y{}^2(1 - r^2)$. If there is no real correlation between the two variables then parts (1) and (2) are estimates of the same universe variance. Now to apply the z-test we must have *unbiased* estimates. There is one degree of freedom in part (1) and $N - 2$ in

TABLE 17

	Variation	*D. of F.*
Regression line	$\sum_x (y - Y_x)^2 f(x) = Nr^2\sigma_y{}^2$	1
Residuals	$\sum_x (Y_x - \bar{y})^2 f(x) = N(1 - r^2)\sigma_y{}^2$	$N - 2$
Totals	$\sum_x (y - \bar{y})^2 f(x) = N\sigma_y{}^2$	$N - 1$

part (2). Consequently we may test the independence of y and x by computing

$$z = \tfrac{1}{2} \log_e \frac{r^2(N - 2)}{1 - r^2} \tag{60}$$

and seeing if it lies beyond the 5% or 1% points in the table for $n_1 = 1$, $n_2 = N - 2$. However, it is conventional to make use of

Fisher's t-distribution. It can be shown (see Problem 10) that the distribution of t is a special case of that for z when $n_1 = 1$, $n_2 = n$, and $z = \frac{1}{2}\log_e t^2$. Therefore,

$$t = r\left\{\frac{N-2}{1-r^2}\right\}^{1/2} \tag{61}$$

is distributed in accord with $F_n(t)$ for $n = N - 2$. In § 11 we observed that the 0.05 *level* of significance for z is the .025 *point*. However, when used as an alternative to t, the 0.05 point of z is also the 0.05 level because the whole distribution of t is equivalent to the positive half of the z-distribution in the sense that, for tests of significance, z ranges from 0 to ∞ whereas t ranges from $-\infty$ to ∞.

Tables are available (Fisher's text, Table V.A.) for applying this test directly from r, giving values of r on four levels of significance represented by $P = .10, .05, .02$, and $.01$, for various values of n. It might prove interesting to compare an entry in this table with the corresponding entries in the z and t tables. For example, when $n = 18$ ($N = 20$) we find from this table that $r = .4438$ lies on the $P = .05$ level, and making the transformation to z by (60) we obtain $z = .7424$ which agrees exactly with the entry in the z-table at the .05 point when $n_1 = 1$, $n_2 = 18$. Finally, when $r = .4438$ in (61) $t = 2.101$ which is the entry in the t-table at the .05 level.

Case II. When $\rho \neq 0$. If the samples are large ($N > 100$) and if ρ is small or only moderately large ($|\rho| < .6$ perhaps) then it is true that r is approximately normally distributed about the value ρ with standard deviation of

$$\sigma_r = (1 - \rho^2)(N - 1)^{-1/2}.$$

It is customary, under these conditions, to attach to an observed value of r a standard error of

$$\sigma_r = (1 - r^2)(N - 1)^{-1/2}$$

and, for a proposed ρ, to refer the computed value of

$$t = \frac{r - \rho}{\sigma_r}$$

to a normal probability scale.

This procedure is invalid, however, if N is small and ρ is large. The distribution of r from small samples is skew and the skewness increases with ρ. This may be understood intuitively by considering the distribution of r's from a universe in which ρ is .9. The range of

possible variation of r above ρ is only .1. But the possible range below ρ is 1.9. Accordingly the sampling distribution of r (N small) from this universe will be sharply skew. An extensive coöperative study of the distribution of r was made in 1917 by Soper and others [23]

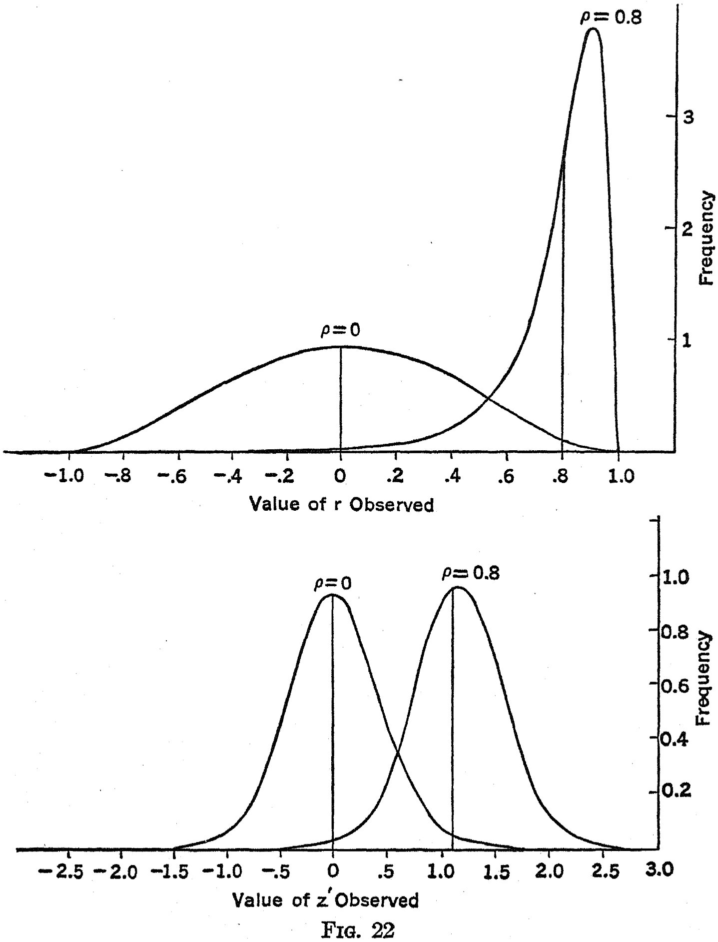


FIG. 22

They succeeded in finding expressions for its moments and on this basis represented the distribution, for various values of N and ρ, by Pearson curves. They also gave an elaborate set of tables of ordinates for values of ρ from 0 to 1 by increments of .1 and for values of r from -1 to $+1$ by increments of .05. The upper panel of Figure 22 (from Fisher's book) shows the r curves for two values

of ρ with $N = 8$, which (presumably) were drawn from the ordinates of these tables. They indicate the rapid departure from normality that may be expected for small samples as ρ approaches high values.

In his study of the sampling distribution of the correlation coefficient Fisher found that it was not desirable to use r as the independent variable and he introduced a transformation which has distinctive merits. He showed that the quantity*

$$z' = \tfrac{1}{2} \log_e \frac{1 + r}{1 - r} \tag{62}$$

is approximately normally distributed and is nearly constant in form as ρ changes. Its mode is always close to ρ. The lower panel of Figure 22 shows the distribution curves for z' corresponding to the r curves in the upper panel. The standard deviation is

$$\sigma_{z'} = (N - 3)^{-1/2} \tag{63}$$

and is practically independent of ρ. The transformation is applicable in the following tests (among others).

(*a*) To test if an observed value of r differs significantly from a proposed theoretical value, ρ.

(*b*) To test if two observed values are significantly different.

The procedure for (*a*) is to calculate†

$$t = (z' - z'')(N - 3)^{1/2} \tag{64}$$

and refer the result to a normal probability scale. For (*b*) the procedure is to find, in accordance with (62), the two values of z', say z'_1 and z'_2, corresponding to the two observed values of r, say r_1 and r_2 from samples of N_1 and N_2, respectively. Then compute $d = z'_1 - z'_2$ and $\sigma_d = \{1/(N_1 - 3) + 1/(N_2 - 3)\}^{1/2}$ and refer

$$t = \frac{d}{\sigma_d}$$

to a normal probability scale.

For numerical examples the student is referred to Fisher's book, §§ 33–35. Tables are also available there to facilitate the computation of z' for an assigned r. One should observe that the z' technique is not applicable to the case of simple tests of significance ($\rho = 0$). In that case Fisher's table of t is available.

* This quantity is not quite the same as the z used for the ratio of two variances and so we use a prime here to distinguish between them.

† In (64), z'' is the value of (62) when r is replaced by ρ.

Three final remarks seem appropriate. (1) In computing an r to be tested it is not desirable to apply Sheppard's corrections to s_x and s_y because they tend to increase the value of r. This also applies in testing for linear regression (§ 15). (2) It has been shown that the z' procedure is applicable in testing the significance of partial correlation coefficients if N in $\sigma_{z'}$, is replaced by $N - k$ where k is the number of secondary subscripts in the coefficient. (3) All of the above procedures are strictly valid only for normal universes. However, there is considerable experimental evidence to indicate that they hold for all practical purposes provided the marginal distributions of one or both variables in the universe are not of the J- or U-shaped types. Of course, in those extreme cases one would naturally hesitate to use r as a measure of association.

References and Notes

1. R. A. Fisher and some other writers use the symbol s^2 to denote an unbiased estimate of σ^2 based on a single sample.
2. A good discussion of the problem of Statistical Estimation is given by
   J. Neyman, *Lectures and Conferences on Mathematical Statistics*, pp. 127–142, 1938. The Graduate School of the United States Department of Agriculture, Washington, D. C.
   See also
   E. S. Keeping, *Note on a Point in the Theory of Sampling.* American Mathematical Monthly, vol. 42, p. 161.
   E. G. Olds, *A Note on the Problem of Estimation.* American Mathematical Monthly, vol. 44, p. 92.
3. Egon Pearson, *The Application of Statistical Methods to Industrial Standardization and Quality Control.* British Standards Institution. 1935.
4. W. A. Shewhart, *Economic Control of Quality of Manufactured Product.* D. Van Nostrand. 1931.
5. A. T. Craig, *A Certain Mean-Value Problem in Statistics.* Bulletin American Mathematical Society, vol. 42, pp. 670–674.
6. "Student," *Probable Error of Mean.* Biometrika, vol. 6, 1908. His real name was W. S. Gosset. Interesting biographical sketches of Mr. Gosset will be found in the Journal of the American Statistical Association, vol. 33, No. 201, March, 1938, pp. 226–228; and the Journal of the Royal Statistical Society, vol. CI, Part I, pp. 248–251.
   For an account of the earlier contributions by Helmert (1876) and Czuber (1891) see *Some Topics in Sampling Theory,* H. L. Rietz. Bulletin American Mathematical Society, April, 1937.
7. R. A. Fisher, *Applications of "Student's" Distribution.* Metron, vol. 5, no. 3, pp. 90–93.
8. R. C. Geary, *The Distribution of "Student's" Ratio for Non-Normal Samples.* Journal of the Royal Statistical Society Supplement, vol. 3, 1936, pp. 178–84.

9. V. Romanovsky, *On the Moments of Standard Deviation and of Correlation Coefficient in Samples from Normal.* Metron, vol. 5, no. 4, pp. 3–46.
10. Paul R. Rider, *A Survey of the Theory of Small Samples.* Annals of Mathematics, (2), vol. 31, no. 4, pp. 577–628.
11. Deming and Birge, *On the Statistical Theory of Errors.* Review of Modern Physics, vol. 6, pp. 119–161. (Reprints available at the Graduate School of the United States Department of Agriculture, Washington, D. C.)
12. Vol. 5, no. 3, pp. 114–120.
13. R. A. Fisher, *Statistical Methods for Research Workers.* Oliver and Boyd. Table IV.
14. Dunham Jackson, *The Theory of Small Samples.* American Mathematical Monthly, vol. 42, 1935, pp. 344–364.
15. John F. Kenney, *A Note on Certain Formulas used in Sampling Theory.* American Mathematical Monthly, vol. 45, pp. 456–458.
16. M. S. Bartlett, *The Effect of Non-Normality on the t-Distribution.* Proc. Camb. Phil. Soc., vol. 31 (1935), pp. 223–231.
17. H. L. Rietz, *Comments on Applications of Recently Developed Theory of Small Samples.* Journal American Statistical Association, vol. 26 (1931), pp. 150–158.
18. P. R. Rider, *A Note on Small Sample Theory.* Journal American Statistical Society, vol. 26, no. 174, pp. 172–174.
19. Table VI in reference 13.
20. G. W. Snedecor, *Statistical Methods.* Collegiate Press, Inc. Ames, Iowa.
21. A. T. Craig, *On the Difference Between Two Sample Variances.* National Mathematics Magazine, vol. 11 (1937), pp. 259–262.
22. A. T. Craig, *On the Independence of Certain Estimates of Variance.* Annals of Mathematical Statistics, vol. 9, pp. 48–55. See also a paper by Irwin entitled *On the Independence of the Constituent Items in the Analysis of Variance.* Journal Royal Statistical Society, Supplement 1–3, 1934–36, pp. 236–251.
23. H. E. Soper *et. al., On the Distribution of the Correlation Coefficient in Small Samples.* Biometrika, vol. 11, pp. 328–413.

Problems

1. Derive the expression for the expected value of s^2 in repeated samples of N independent observations from an arbitrary universe. Explain the use of this expression in estimating the variance of a universe.
2. In a certain observed distribution, $N = 20$, $\bar{x} = 42$, $s = 5$. Test the hypothesis that this distribution is a random sample from a normal universe with mean of 50.
3. In a certain test, one section of 20 students had an average score of 40 with a standard deviation of 5. Another section of 25 had an average of 46 with standard deviation of 4. Does this indicate a significant difference in the two groups? What assumptions do you make in applying the test?

4. In an experiment in industrial psychology a job was performed by one group of 30 workmen according to Method I and by a second group of 40 according to Method II. (The groups were independent and equally efficient.) Are the following distributions of the time (in seconds) taken such as to justify the conclusion that Method I is the speedier of the two? Use the difference between the means as a criterion of judgment.

Time	*I*	*II*
50	1	0
51	3	1
52	5	2
53	4	5
54	7	8
55	5	9
56	3	6
57	1	3
58	1	3
59	0	1
60	0	2
Totals	30	40

5. From the separate distribution functions of $\bar{x}$ and s derive the distribution of "Student's" z, and from that obtain the function $F_n(t)$.
6. Prove that $F_n(t)$ is asymptotically normally distributed.
7. Derive Fisher's z-distribution, $G(z)$.
8. (*Mills' text, revised.*) Manufacturing industries were classified into those producing perishable, semi-durable, and durable goods. An average of changes occurring between 1929 and 1933 in the selling prices of the products of each of these categories was computed giving the index numbers shown in the y_x column of the following table.

Class of industry, x	*Number of industries,* N_x	*Means,* $\bar{y}_x$	*Computations*
Producing perishable goods	34	69.81	$b - 1 = 2, \quad N - b = 82$
Producing semi-durable goods	26	66.41	$v_1 = 2{,}161.8800$
Producing durable goods	25	78.96	$v_2 = 15{,}564.9040$
All industries	85		$V = 17{,}726.7840$

Compute F and test the null hypothesis that there was no real difference in the price movements of the three different classes of industry for the years 1929–1933.

9. Prove that $$\sum_{x=1}^{2} N_x(\bar{y}_x - \bar{y})^2 = \frac{N_1 N_2}{N_1 + N_2}(\bar{y}_1 - \bar{y}_2)^2.$$

10. Prove that the test for significance between two means is a special case of the test for significant variation in sub-sets of means by showing that (56) of § 14 reduces, when $b = 2$, to

$$\frac{\bar{y}_1 - \bar{y}_2}{\hat{\sigma}}\left\{\frac{N_1 N_2}{N_1 + N_2}\right\}^{1/2}$$

where $\hat{\sigma}$ is an unbiased estimate of σ and t is distributed in accord with $F_n(t)$ for $n = N_1 + N_2 - 2$.

The following three problems are from Fisher's book.

11. For the twenty years 1885–1904, the mean wheat yield of Eastern England was found to be correlated with the autumn rainfall; the correlation was found to be $-.629$. Is this value significant?

12. In a sample of $N = 25$ pairs of parent and child the correlation in a certain character was found to be .60. Is this value consistent with the view that the true correlation in that character was .46?

13. Of two samples the first, of 20 pairs, gives a correlation of .6, the second, of 25 pairs, gives a correlation .8. Are these values significantly different?

CHAPTER VIII

A. THE χ^2 DISTRIBUTION AND APPLICATIONS

1. The Multinomial Law.[1] The general term of the multinomial expansion for k mutually exclusive categories sets the stage for a presentation of χ^2 which provides an insight into the probability theory of this important quantity and its usefulness in the testing of hypotheses. So we begin with a preliminary treatment of the multinomial law.

Consider an event that is characterized by a variable v which can take on one of k values, $v_1, v_2, \cdots v_k$. Let the probability that v_i occurs be p_i, where $\sum_1^k p_i = 1$. Then in N independent trials, the probability that v_1 occurs m_1 times, v_2 occurs m_2 times, and so on, in a *specified order* (whatever it may be) is

$$p_1^{m_1} p_2^{m_2} \cdots p_k^{m_k}$$

where $\sum_1^k m_i = N$, the m's being positive integers or zero. The number of ways in which the order can be specified is the number of permutations possible among N objects of which m_1 are of type T_1, m_2 of type $T_2, \cdots m_k$ of type T_k. Let this number be denoted by $p[m_i]$. Then we have

$$p[m_i] = \frac{N\,!}{m_1\,!\,m_2\,!\cdots m_k\,!}.$$

Therefore, the probability that m_1 of the variates take the value v_1, m_2 the value v_2, and so on, *regardless of order* is

$$f(m_1, m_2, \cdots m_k) = p[m_i] p_1^{m_1} p_2^{m_2} \cdots p_k^{m_k} \tag{1}$$

which is the general term of the expansion of the multinomial

$$(p_1 + p_2 + \cdots + p_k)^N.$$

The law of repeated trials, for a simple dichotomy, given in Chapter I, is a special case of this law. Thus if $k = 2$, the right member

of (1) reduces to

$$C(N, r)p^r q^{N-r} \tag{2}$$

where

$r = m_1,\ N - r = m_2,\ p = p_1,\ q = 1 - p_1 = p_2,\ C(N, r) = N!/m_1!\,m_2!$.

If v is the number of spots appearing on the top face in a throw of a die, then v will take on one of the values 1, 2, 3, 4, 5, 6, and the probability of throwing exactly r aces (say) in N throws of the die is

$$C(N, r)(\tfrac{1}{6})^r(\tfrac{5}{6})^{N-r}.$$

We recall that (2) is the general term of the expansion of the binomial $(q + p)^N$. By using Stirling's approximation for factorials, we can derive an approximation for (1) which will bear to the multinomial law a relation analogous to that which the normal curve bears to the binomial. With this objective in mind, assume that every m_i is sufficiently large for $m_i!$ to be replaced by its Stirling approximation. Making these replacements (1) becomes, after some algebraic rearrangement,

$$f(m_1, m_2, \cdots m_k) = \frac{\prod_{i=1}^{k} (Np_i/m_i)^{m_i+1/2}}{(2\pi N)^{(k-1)/2}(p_1p_2 \cdots p_k)^{1/2}}. \tag{3}$$

Next introduce the transformation

$$t_i = \frac{m_i - Np_i}{\sigma_i}, \tag{4}$$

σ_i^2 being $Np_i(1 - p_i)$. Under this transformation (3) becomes

$$(2\pi N)^{(k-1)/2}(p_1p_2 \cdots p_k)^{1/2} f = \prod_{i=1}^{k} \left(1 + \frac{\sigma_i t_i}{Np_i}\right)^{-Np_i - \sigma_i t_i - 1/2}$$

Then

$$\log \text{L.M.} = \sum_{1}^{k} (-Np_i - \sigma_i t_i - \tfrac{1}{2}) \log\left(1 + \frac{\sigma_i t_i}{Np_i}\right)$$

where L.M. denotes the left-hand member of the preceding equation. Upon expanding the logarithm in a power series and collecting the results according to descending powers of N, we obtain

$$\log \text{L.M.} = -\sum_{1}^{k}\left(\sigma_i t_i + \frac{\sigma_i^2 t_i^2}{2Np_i} + \text{terms of lower order}\right).$$

Let each m_i in $\sum m_i = N$ be transformed in accordance with (4). The result is

$$\sum_1^k \sigma_i t_i + N\sum_1^k p_i = N,$$

whence it follows that $\sum \sigma_i t_i = 0$ since $\sum p_i = 1$. Therefore, remembering the value of σ_i^2, (3) may be written

$$f(m_1, m_2, \cdots m_k) = (2\pi N)^{(1-k)/2}(p_1 p_2 \cdots p_k)^{-1/2} e^{-\frac{1}{2}\sum_1^k t_i^2(1-p_i)}.$$

The form of the exponent of e suggests the substitution of a new variable $x_i = t_i(1 - p_i)^{1/2}$ in place of t_i. Upon making this substitution we have

$$(5) \qquad f(m_1, m_2, \cdots m_k) = (2\pi N)^{(1-k)/2} (p_1 p_2 \cdots p_k)^{-1/2} e^{-\frac{1}{2}\sum_1^k x_i^2}$$

where $x_i = (m_i - Np_i)(Np_i)^{-1/2}$ and Np_i is the mean or expected value of m_i.

Now, following Wilks,[2] the x's are independent except for the single linear restriction $\sum_1^k (Np_i)^{1/2} x_i = 0$. Let R be the region in the x-space subject to the linear restriction just given corresponding to any region R_m in the m-space. Since the m's are always integers, the change in x_i corresponding to a change of unity in m_i is $(Np_i)^{-1/2} = \Delta x_i$. Treating $k - 1$ of the x's, say $x_1, x_2, \cdots x_{k-1}$ as the independent variables, and using an extension of the fundamental theorem on the existence of a definite integral (Riemann), we have

$$(6) \qquad \lim_{N\to\infty} \sum_R f(x_1, x_2, \cdots x_k)\, \Delta x_1 \Delta x_2 \cdots \Delta x_{k-1} = \frac{1}{(2\pi)^{(k-1)/2}(p_k)^{1/2}} \int_R e^{-\frac{1}{2}\sum_1^k x_i^2}\, dx$$

where for a given N, $\sum_R$ denotes the summation over all points in the region R corresponding to those in R_m for which $f(m_1, m_2, \cdots m_k)$ is defined. The integral is k-dimensional, and $dx = dx_1\, dx_2 \cdots dx_{k-1}$.

2. The χ^2 Distribution. The quantity

$$(7) \qquad \sum_1^k x_i^2 = \chi^2$$

is used as an index of the extent to which the set of m's taken as a whole cluster about their respective expected values. Later on we will explain the practical import of this index. For the present we

confine our attention to the purely mathematical problem of finding the distribution function of χ^2. First, we consider the problem of finding the distribution function of χ. To this end we observe that, corresponding to different values of χ, (7) defines a set of k-dimensional hyperspheres all having their centers at the origin of the x_i-axes and no two intersecting. Now we can obtain the distribution of χ by determining the value of the integral in (6) when R consists of the region bounded by the concentric hyperspheres

$$(8) \qquad \sum_1^k x_i^2 = \chi^2 \quad \text{and} \quad \sum_1^k x_i^2 = (\chi + d\chi)^2$$

subject to the condition that

$$(9) \qquad \sum_1^k (Np_i)^{1/2} x_i = 0.$$

Since this last equation is a hyperplane through the common center of the hyperspheres, the region R is therefore a " shell " of a $k - 1$ hypersphere. Within this shell

$$e^{-\frac{1}{2}\sum_1^k x_i^2} = e^{-\frac{1}{2}\chi^2}$$

to within terms of order $d\chi$.

Now it can be shown that the volume V of an s-dimensional hypersphere of radius r is

$$V = Cr^s$$

where C is independent of r. The volume between two concentric hyperspheres of radii r and $r + dr$ is therefore approximately

$$(10) \qquad dV = C' r^{s-1}\, dr.$$

Returning to the χ problem, it is clear from (10) that if the region bounded by the hyperspheres in (8), subject to the restriction given by (9), is chosen as the element of volume, then the probability that $\left\{\sum_1^k x_i^2\right\}^{1/2}$ will lie in the interval from χ to $\chi + d\chi$ is

$$(11) \qquad df = Ke^{-\frac{1}{2}\chi^2} \chi^{k-2}\, d\chi.$$

Here K is independent of χ and can be determined by the condition $\int_0^\infty df = 1$. Using the Gamma function, we find

$$\frac{1}{K} = 2^{(k-3)/2}\, \Gamma\left(\frac{k-1}{2}\right).$$

The distribution of χ^2 is thus given by

$$(12)\qquad T_{k-1}(\chi^2)\,d(\chi^2) = \frac{(\chi^2)^{(k-3)/2}e^{-\frac{1}{2}\chi^2}}{2^{(k-1)/2}\Gamma\left(\frac{k-1}{2}\right)}\,d(\chi^2)\,.$$

The number $k - 1$ is the number of degrees of freedom which is the number of x's which are independent in (6).

3. Tables. The probability of obtaining a sample of x's for which $\sum x_i^2$ is greater than an assigned χ^2, say χ_0^2, is given by

$$(13)\qquad P(\chi^2 > \chi_0^2) = \int_{\chi_0^2}^{\infty} T_{k-1}(\chi^2)\,d(\chi^2).$$

The symbol on the left in (13) may be abbreviated to P when there is no ambiguity. It is obvious that χ^2 is never negative and may vary from 0 (when there is no difference between the observed and expected frequencies) to very large values. As χ^2 increases from 0 to ∞, the probability P given by (13) decreases from 1 to 0. The student will recognize T_{k-1} (χ^2) as a Pearson Type III curve and the integral in (13) as essentially an incomplete Gamma function. Values of P can be found in Pearson's *Tables*[3] and we have included in the Appendix (see Table III) a short table, from Fisher's book,[4] giving values of χ^2 corresponding to specially selected values of P. In our table, $n = k - 1$.

For fairly large values of k, $(2\chi^2)^{1/2}$ is approximately normally distributed about a mean $(2k - 1)^{1/2}$ with unit standard deviation. Therefore, one may refer

$$t = (2\chi^2)^{1/2} - (2k - 1)^{1/2}$$

to a normal probability scale when $k > 30$.

4. Applications. The χ^2-test was designed by its originator, Karl Pearson,[5] as a criterion for testing hypotheses about frequency distributions. These hypotheses may be classified into two types which we will call *simple* and *composite*. We are making an explicit distinction between them and considering them separately to avoid certain misunderstandings which have sometimes occurred, in the past, in the applications of the test. To be more specific, there has been, as a result of confounding hypotheses to be tested, some controversy over the appropriate number of degrees of freedom to use in entering the tables for $P(\chi^2 > \chi_0^2)$.

Simple Hypothesis. Under this heading we will consider those cases in which the theoretical frequencies are known *a priori*, that is, when they are not inferred in any way from the sample.

Suppose that we have a set of k observed frequencies

$$m_1 + m_2 + \cdots + m_k = N$$

constituting a sample from a hypothetical universe (supposedly infinite) in which the relative frequencies in the k categories are known to be $p_1, p_2, \cdots, p_k$, respectively, where $p_i = \tilde{m}_i/N$. Then, corresponding to the observed frequencies, we have a set of k theoretical frequencies such that

$$\tilde{m}_1 + \tilde{m}_2 + \cdots + \tilde{m}_k = N.$$

An example would be, for the m's, the frequency of heads obtained in tossing N coins k times, and, for the $\tilde{m}$'s, the corresponding theoretical frequencies given by the terms in the expansion of the binomial $N(\frac{1}{2} + \frac{1}{2})^k$. In comparing the observed and theoretical frequencies a question quite naturally arises as to whether the aggregate discrepancy between them could be explained on the basis of chance fluctuations under the hypothesis that $\frac{1}{2}$ is the probability of success in each trial. More generally, we are interested in such a question as the following. On the hypothesis that an observed distribution is a random sample from a proposed universe, what is the probability that, taken as a whole, the discrepancy between theory and observation would yield a value of χ^2 as large as, or larger than, the value obtained. The hypothesis is to be rejected whenever the probability is considered "small."

If we let $x_i = (m_i - \tilde{m}_i)/\sqrt{m_i}$ it is clear that the x's are subject to the linear homogeneous restriction given by (8) with $n = k - 1$ degrees of freedom because, if $k - 1$ of the x's are fixed, the kth is determined. In the case of a simple hypothesis, then, Fisher's table of P is to be entered with $n = k - 1$.

With regard to levels of significance, Fisher[4] says:

> In preparing this table we have borne in mind that in practice we do not want to know the exact value of P for any observed χ^2, but, in the first place, whether or not the observed value is open to suspicion. If P is between .1 and .9 there is certainly no reason to suspect the hypothesis tested. If it is below .02 it is strongly indicated that the hypothesis fails to account for the whole of the facts. We shall not often be astray if we draw a conventional line at .05, and consider that higher values of χ^2 indicate a real discrepancy.

Composite Hypothesis. In the majority of practical cases, the frequencies are not known *a priori* and must be estimated from the sample. Thus, in a graduation by means of the normal curve the theoretical frequencies are obtained by imposing the conditions that the universe has the same mean and standard deviation as the sample. The χ^2-test can be accurately applied only if allowance is made for the number of parameters which are determined from the sample in reconstructing the universe. Suppose there are q parameters in the function representing the universe and these are to be determined from the sample by the principle of moments. Since any moment is a linear function of the frequencies (it will be remembered that the frequencies are the variables in this discussion), the determination of the q parameters involves q linear restrictions. We have seen in § 2 that the restriction imposed by (9) reduced our problem from a space of k dimensions to a space of $k - 1$ dimensions. Quite analogously, q additional linear restrictions reduce the space to $k - 1 - q$ dimensions. Accordingly,* in testing divergence from a universe specified by a function $f(v, a, b, c, \cdots)$ where v is the variable of the distribution and $a, b, c, \cdots$ are q disposable parameters which are to be estimated from the sample, the number of degrees of freedom with which to enter the tables of P is $n = k - 1 - q$.

The following two conditions should be fulfilled in applying the χ^2-test (for both simple and composite hypotheses).

1. No class should contain very few items because, in the derivation of (11), it was assumed that m_i was sufficiently large to replace $m_i\,!$ by its Stirling approximation.

2. The number of classes should not be very large since it can be shown, by expanding the integrand in (13) into a power series, that $P \rightarrow 1$ as $k \rightarrow \infty$.

We shall interpret, somewhat arbitrarily, these conditions to mean that P cannot be guaranteed when $m < 5$ and $k > 20$. To satisfy the first condition, it is customary to lump together the small frequencies at the ends of the distribution.

Example 1. Twelve dice were thrown 4096 times; only a throw of six was counted a success. The expected frequencies are given by $4096(\frac{1}{6} + \frac{5}{6})^{12}$. How improbable, taken as a whole, is the observed distribution shown in Table 18?

* Strictly speaking, the determination of the parameters by the method of moments does not lead to a system of equations which are exactly analogous to (9).

TABLE 18

Number of Successes	*Observed Frequency*	*Theoretical Frequency*	$(m - \tilde{m})^2$	$\frac{(m - \tilde{m})^2}{m}$
0	447	459	144	.3137
1	1145	1103	1764	1.5993
2	1181	1213	1024	.8442
3	796	809	169	.2089
4	380	364	256	.7033
5	115	116	1	.0086
6	24	27	9	.3333
7 and over	8	5	9	1.8000
Totals	4096	4096		$\chi^2 = 5.8113$

Entering Table III (see Appendix) with $n = 8 - 1 = 7$, and interpolating for the value of P corresponding to the observed value of $\chi^2 = 5.8113$, we find $P = .56$. Hence there is no reason to reject the hypothesis that the underlying chance of a "success" is $p = \frac{1}{6}$. That is, there is no reason to suspect that the dice were biased.

Example 2. An observed distribution was graduated by means of the normal curve (see Part I, p. 123) with the results shown in Table 19. Test the hypothesis that the observed distribution was a sample from a normal universe with mean and standard deviation equal respectively to those of the sample.

TABLE 19

Central Values	*Observed Frequency*	*Theoretical Frequency*
29.5	15 { 1	17.2 { 2.5
33.5	14	14.7
37.5	56	60.2
41.5	172	155.4
45.5	245	252.6
49.5	263	258.8
53.5	156	167.2
57.5	67	68.0
61.5	26 { 23	20.6 { 17.5
65.5	3	3.1
Totals	1000	1000.0

It is found that $\chi^2 = 4.82$. After pooling the end frequencies, as shown, $k = 8$. So entering Table III for $n = 8 - 1 - 2 = 5$, we find that $P > .4$. Hence the χ^2-test does not reject the hypothesis.

For applications of the χ^2-test to *contingency tables* the reader is referred to Fisher's book.

B. STATISTICAL INFERENCE

5. Induction versus Deduction. To contrast the inductive problems, which we are about to consider, with deductive problems, we shall review briefly a deductive type of argument which we have previously discussed. Suppose $D(t)$ is the distribution function of a statistic t computed from a sample from a universe specified with respect to functional form and parameters. Then $\int_{-\infty}^{\delta} D(t)\,dt$ gives the probability that an observed value of t will not exceed an assigned value of δ. Thus in Chapter VI we learned that the means of samples cluster about the mean of the universe, and Theorem X of that chapter gave us the probability that a sample mean would have a numerical value within δ of the mean of the universe. This is a deductive argument. Presently we shall consider certain inverse problems which arise in arguing from samples and their statistics back to universes and their parameters. First, however, we shall examine Bayes' Theorem. The following quotation from R. A. Fisher[6] will serve as a setting for our consideration of this theorem.

Thomas Bayes' paper of 1763 was the first attempt known to us to rationalize the process of inductive reasoning. From time immemorial, of course, men had reasoned inductively; sometimes, no doubt, well, and sometimes badly, but the uncertainty of all such inferences from the particular to the general had seemed to cast a logical doubt on the whole process. By the middle of the eighteenth century, however, experimental science had taken its first strides, and all the learned world was conscious of the effort to enlarge knowledge by experiment, or by carefully planned observation. To such an age the limitations of a purely deductive logic were intolerable. Yet it seemed that mathematicians were willing to admit the cogency only of purely deductive reasoning. From an exact hypothesis, well defined in every detail, they were prepared to reason with precision as to its various particular consequences. But, faced with a finite, though representative, sample of observations, they could make no rigorous statements about the population from which the sample had been drawn.

Bayes perceived the fundamental importance of this problem and framed an axiom, which, if its truth were granted, would suffice to bring this large class of inductive inferences within the domain of the theory of probability; so that, after a sample had been observed, statements about the population could be made, uncertain inferences, indeed, but having the well-defined type of uncertainty characteristic of statements of probability. Bayes' technique in this feat is ingenious. His predecessors had supplied adequate methods, given a well-defined population, for stating the probability that any particular type of population might have given rise to it. He imagines, in effect, that the possible types of population have themselves been drawn, as samples, from a super-population, and his axiom defines this super-population with exactitude. His

problem thus becomes a purely deductive one to which familiar methods were applicable.

6. Bayes' Theorem. To derive Bayes' theorem, consider a bivariate universe of discrete variables in which x takes the values $x_1, x_2, \cdots, x_n$, and y the values $y_1, y_2, \cdots, y_m$. Let $P(x_i, y_j)$ represent the probability for the joint occurrence of (x_i, y_j). Let $P(y_j \mid x_i)$ be the probability that y takes the value y_j when it is known that x has taken the value x_i. Then

$$P(y_j \mid x_i) = \frac{P(x_i, y_j)}{g(x_i)}, \tag{14}$$

where $g(x_i) = \sum_{j=1}^{m} P(x_i, y_j)$ is the marginal distribution of x in the bivariate universe and represents the *a priori* probability that x takes the value x_i. Let us write (14) in the form

$$P(x_i, y_j) = g(x_i)P(y_j \mid x_i). \tag{15}$$

By a similar argument we may write

$$P(x_i, y_j) = h(y_j)P(x_i \mid y_j), \tag{16}$$

where $h(y_j) = \sum_{i=1}^{n} P(x_i, y_j)$ is the marginal distribution of y, and $P(x_i \mid y_j)$ is the probability that $x = x_i$ when it is known that $y = y_j$. It is clear from preceeding relations that

$$h(y_j) = \sum_{i=1}^{n} g(x_i)P(y_j \mid x_i). \tag{17}$$

Since $P(x_i, y_j)$ means exactly the same thing in (15) and (16) we may equate their right members and solve for $P(x_i \mid y_j)$. The result is

$$P(x_i \mid y_j) = \frac{g(x_i)P(y_j \mid x_i)}{h(y_j)}. \tag{18}$$

This is Bayes' theorem and it may be stated as follows.

Bayes' Theorem. *The probability that $x = x_i$ when $y = y_j$ is equal to the product of the probabilities that $x = x_i$, and that $y = y_j$ when $x = x_i$, divided by the probability that $y = y_j$.*

The theorem is usually expressed symbolically in the somewhat different form to which it reduces when (17) is substituted for the

denominator of (18). This form is

$$P(x_i \mid y_j) = \frac{g(x_i)P(y_j \mid x_i)}{\sum_{i=1}^{n} g(x_i)P(y_j \mid x_i)}. \tag{19}$$

To connect Bayes' theorem with *a posteriori*, or *inverse*, probability suppose in (19) that the x's denote certain initial situations and the y's denote events subsequently observed. The *a priori* probability for the existence (occurrence) of the initial situation characterized by x_i is $g(x_i)$. $P(y_j \mid x_i)$ is the *a priori* probability that y_j will occur when x_i exists. Then (19) gives the *a posteriori* probability that the ith initial situation has produced the observed event specified by y_j.

The following examples will clarify the theorem and serve to focus attention on its weakness. The first example, a somewhat artificial one, is designed to illustrate a situation where the existence probabilities $g(x_i)$ are equal. The second will describe a situation when nothing is known about them.

Example 3. (*Molina*[7]) During his sophomore year Tom Smith played on both the baseball and football teams; we have been informed that he broke his ankle in one of the games; what are the *a posteriori* probabilities in favor of baseball and football, respectively, as the baneful cause of the accident? Evidently the answer depends on the number of baseball and football games played during their respective seasons and also on the likelihood of a man breaking an ankle in one or the other of these two sports. As a concrete case assume that:

(*a*) At Smith's college an equal number of baseball and football games are played per season;

(*b*) Statistical records indicate that if a student participates in a baseball game the probability is $\frac{2}{100}$ that he will break an ankle and that, likewise, the probability is $\frac{7}{100}$ for the same contingency in a football game.

Solution. Associate x_1 and x_2 with the admissible causes, baseball and football, respectively. Associate y_1 with the accident. From condition (*a*) of the problem, the existence probability for baseball is $g(x_1) = \frac{1}{2}$. Also $P(y_1 \mid x_1) = \frac{2}{100}$, and $P(y_1 \mid x_2) = \frac{7}{100}$. From (19), then, the *a posteriori* probability for baseball is

$$P(x_1 \mid y_1) = \frac{\frac{1}{2}\cdot\frac{2}{100}}{\frac{1}{2}\cdot\frac{2}{100} + \frac{1}{2}\cdot\frac{7}{100}} = \frac{2}{9}.$$

It follows that the *a posteriori* probability in favor of football is $\frac{7}{9}$.

Example 4. An urn contains five balls, black, white, or both kinds. Of three balls drawn together and at random (each ball within the urn is equally likely to

be drawn), two are black and one is white. What is the probability that the urn contains three black and two white balls?

Solution. Associate $x_1, x_2 \cdots, x_6$, with the possible compositions of the urn before the drawing was made, namely, $0B, 5W; 1B, 4W; \cdots; 5B, 0W$. Associate y_1, y_2, y_3, y_4, with the possible compositions in the drawing of three balls, namely, $0B, 3W$; $1B, 2W$; $2B, 1W$; $3B, 0W$. The composition corresponding to y_3 was obtained and we seek the probability that it came from an urn with composition specified by x_4. That is, we seek $P(x_4 \mid y_3)$. Clearly,

$$P(y_3 \mid x_4) = \frac{C(3, 2)C(2, 1)}{C(5, 3)} = \frac{3}{5},$$

so from (19) we have

$$P(x_4 \mid y_3) = \frac{g(x_4)\dfrac{3}{5}}{\displaystyle\sum_{i=1}^{6} g(x_i)\frac{C(i, 1)C(5-i, 2)}{C(5, 3)}}, \tag{20}$$

it being understood, of course, that $C(n, r) = 0$ when $n < r$.

Since the values of $g(x_i)$ are unknown the problem does not have a unique solution. Moreover, if they were known we would be back in the domain of deductive probabilities again since all the probabilities in the right-hand member of (20) would then be known *a priori*. It is only when $g(x_i)$ are unknown that we are properly in the domain of *a posteriori* probability. In practical problems the $g(x_i)$ are scarcely ever known.

Bayes realized this and argued that the x's may be considered equally probable unless we have some reason to think they are not. Under this "doctrine of insufficient reason," the x's are assumed to have equal existence probabilities. In this case, $g(x_i) =$ constant and would cancel in (19), thus permitting a definite solution in (20). It appears that Bayes had serious doubts about this "doctrine" for he withheld his entire treatise from publication until his doubts should be resolved, and it was only after his death that his paper was published by friends. Laplace, however, was less cautious, and he incorporated the doubtful theorem into his *Théorie Analytique des Probabilités*. Robed in the authority of Laplace it went unquestioned for a long time. Boole was the first, in 1854, to criticize the assumption of "the equal distribution of our knowledge, or rather of our ignorance" and "the assigning to different states of things of which we know nothing, equal degrees of probability." Today, it is well known that the assumption of constant existence probabilities may lead to mathematical contradictions. This may clearly be seen

in the analogue to (19) for continuous variables. The following illustration of such a contradiction is cited by Wilks (*loc. cit.*).

Let θ be a parameter characterizing the universe and t a statistic from the sample. Then the analogue to (19) for the continuous case is

$$F(\theta \mid t)\,d\theta = \frac{g(\theta)f(t \mid \theta)\,d\theta\,dt}{dt\int g(\theta)f(t \mid \theta)\,d\theta}. \tag{21}$$

Now, if according to the " doctrine of insufficient reason " we may assume $g(\theta)$ to be constant, (21) reduces to

$$F(\theta \mid t)\,d\theta = \frac{f(t \mid \theta)\,d\theta}{\int f(t \mid \theta)\,d\theta}. \tag{22}$$

But by the very nature of this " doctrine " there is no more reason to assume the *a priori* probability function of θ to be constant than there is to assume the *a priori* probability distribution of some function of θ, say θ^2, to be constant. The *a priori* distribution of $\theta^2 = z$ is $g(\sqrt{z})/2\sqrt{z}$. If $g(\sqrt{z})/2\sqrt{z}$ is constant, then

$$F(\theta \mid t)\,d\theta = \frac{\theta f(t \mid \theta)\,d\theta}{\int \theta f(t \mid \theta)\,d\theta}$$

which is certainly inconsistent with (22).

In arguing from a sample to the universe, any inference must be attended with some degree of uncertainty. But uncertainty should not be confused with lack of rigor. As we shall see, statements can be made about population parameters, subject to risks of being wrong, where the error is precisely expressed in terms of probability theory. In other words, the nature and degree of the uncertainty can be rigorously expressed. This can be accomplished without any assumptions regarding the *a priori* existence probabilities.

7. Probable Error. The following concise exposition of the various usages of the term " probable error " is due to Professor A. T. Craig.

There are in the literature three conceptions of the probable error. If, purely for convenience of language, we refer to the probable error of the mean, these conceptions can be stated as follows: (i) The probable error of the mean is that deviation, extended on both sides of the mean of the *population*, such that $\frac{1}{2}$ is the probability that the mean of a *sample* will fall in this interval; (ii) The probable error of a mean is that deviation, extended on both sides of the mean of a

sample, such that $\frac{1}{2}$ is the probability that the mean of the *population* lies in this interval; (iii) The probable error of the mean is that deviation, extended on both sides of the mean of a *sample,* such that $\frac{1}{2}$ is the probability that the mean of another *sample* will fall in this interval. Conception (i) leads without difficulty to the usual formula $.6745(\sigma/\sqrt{N})$ for the probable error of the mean. This formula is rigorously correct for samples of any size drawn from a normal population and is valid for large samples drawn from any population with finite variance. On the other hand, the formula cannot be established under conception (ii) without further assumptions. If, before the sample is drawn, it is assumed, in the absence of any knowledge concerning the distribution of possible values of the mean of the population, that the existence distribution is constant, then the formula admits mathematical proof. But this assumption is essentially the same assumption as that made in applying Bayes' Theorem to problems of probability *a posteriori.*

The modern method of expressing the reliability of a statistical estimate of a population parameter in terms of fiducial limits seems likely to replace the traditional but often misleading mode of expression involving probable error. The rest of the chapter is devoted to this recent advance in statistical inference.

8. Fiducial Theory. The material of this section is reproduced from a recent paper on this subject by Rietz.[8]

In explaining the meaning of the probable error of a statistic, one of the usual types of definition is essentially the following: *The probable error of a statistic, t, is a positive number, E_t, such that the chances are even that the population parameter of which t is an estimate from the sample, will fall within the interval $t - E_t$ to $t + E_t$.*

This definition contains an inference about the values of a population parameter on the basis of information obtained from a random sample drawn from the population.

Formulas for E_t, in terms of observed data, when t may represent any one of a considerable number of statistics, say an arithmetic mean or a correlation coefficient, are usually listed for convenient application in numerous textbooks for teaching courses in statistics.

Under the definition stated above, it is noteworthy that these formulas depend on a fundamental assumption whose validity has long been in doubt. The assumption in question is to the effect that initially, that is, before our drawings of a sample are made, in our

lack of knowledge about the distribution of possible values of an unknown parameter, say of θ, we may assume the existence distribution of θ to be constant.

The invalidity of this assumption in many applied problems of statistical interest may be seen clearly in cases of a continuous distribution function with a derivative. Suppose that our initial assumptions relating to a parameter θ were such that θ would initially be distributed in accord with a continuous frequency function, $g(\theta)$, which has a derivative at each point within its possible range on θ, say from $\theta = \alpha$ to $\theta = \beta$. Next, suppose $g(\theta)$ were restricted to be constant throughout the range of θ. Then it is well known that the distribution of a simple non-linear function of θ would not be constant. For example, the distribution of $z = \theta^n$ ($n \neq 1$, θ real and non-negative) would not be constant, but would be distributed in accord with a frequency function $(1/n)z^{(1-n)/n}$. But if θ is a population parameter, it seems fairly obvious that the logical character of our theory should usually, if not always, be such as to enable us to use a power of θ as a parameter if we found it convenient to do so.

The preceding introduction is designed to lead up to the important fact that, although in the usual statistical inquiry by sample, the true value of the population parameter θ is unknown and remains unknown, there are cases in which precise statements can be made in terms of probabilities about the bounds within which a parameter θ lies without making an assumption about the initial distribution of the possible values of θ. It has been only about seven years since R. A. Fisher initiated some important ideas in this connection to which interesting contributions have been made by several mathematical statisticians.[9–11]

For simplicity, consider a case of a single parameter, θ, in which we know the frequency function of the statistic, t, to be given by an integrable function

$$y_t = f(t, \theta), \tag{23}$$

where the values of t obtained from observation may be assumed to be good estimates of θ. Suppose we know (23) in such form that it is possible to calculate a table of values of the probabilities that the statistic, t, will fall into an assigned interval selected on a possible range (a, b) for any assigned value of θ within the possible range (α, β) of θ.

Next, for illustration, select a positive number ϵ, say $\epsilon = .005$,

on which to base a certain level of confidence about values of θ to be expressed in terms of probabilities.

As our main problem may be clarified by a geometrical representation, conceive of corresponding values of t and θ obtained in an extensive statistical experiment as represented by rectangular coördinates within the rectangle bounded by lines $t = a$, $t = b$, $\theta = \alpha$, $\theta = \beta$. (Fig. 23.)

Consider an arbitrary assignment for θ, say that $\theta = \theta'$ is the true value of θ. This gives the line AB (Fig. 23). Since the distribution of the statistic t is assumed to be known for each assigned value of θ, we may locate on the line AB two points, t_1 and t_2 ($t_1 \leqq t_2$) such that ϵ is equal to the probability that t obtained from a random sample will yield a value of t less than or equal to t_1, and similarly ϵ is the probability that such a sample will yield a value greater than or equal to t_2. Then we have an interval on AB from t_1 to t_2 such that $1 - 2\epsilon$ is the probability that the random sample will yield a value within this interval.

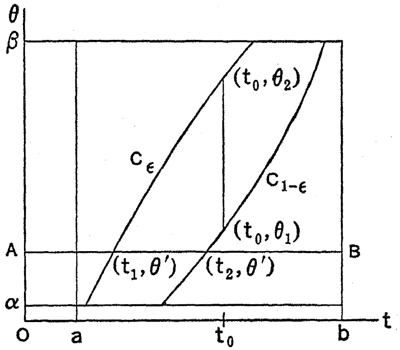


FIG. 23

More formally stated, we may introduce a function $F(t, \theta)$ defined as the definite integral of $f(t, \theta)$ in (23) from $t = a$ to t. That is,

$$F(t, \theta) = \int_a^t f(t, \theta)\, dt,$$

for any arbitrarily assigned real value of θ on its range from α to β. Then

$$F(a, \theta) = 0, \qquad F(b, \theta) = 1, \qquad F(t_1, \theta') = \epsilon, \qquad F(t_2, \theta') = 1 - \epsilon,$$
$$(0 < \epsilon < 1).$$

By considering all possible assignments of θ, in its possible range (α, β), the locus of our set of lower values of t, illustrated by t on the line AB, will give a continuous curve which we mark with C_ϵ in Figure 23, the subscript ϵ being used to remind us that ϵ is the probability that a random value of t for $\theta = \theta'$ will fall below or at t_1.

Similarly, our set of upper values of t, illustrated by t_2 on AB, give a curve which we mark with $C_{1-\epsilon}$.

If t is a good estimate of θ, its value usually, if not always, increases with θ for all possible values. Thus, we shall restrict our further considerations to cases in which we may assume that t increases as θ increases and vice versa. More precisely we are concerned with one-valued monotone increasing functions represented by the two curves marked C_ϵ and $C_{1-\epsilon}$. The region bounded by these two curves and the lines $\theta = \alpha$ and $\theta = \beta$ has been called by Neyman the *confidence belt* with *confidence coefficient* equal to $1 - 2\epsilon$.

Next, consider the set of points, (t, θ), that would be obtained in Figure 23 in carrying out an extensive statistical experiment for which we seek a degree of accuracy in the long run, indicated by the value we assign to ϵ. Then it is fairly obvious that the confidence belt is so constructed that $1 - 2\epsilon$ is the expected relative frequency with which points, (t, θ), will lie inside the confidence belt, and 2ϵ is the expected relative frequency with which such points will lie outside the confidence belt or on its boundary, whatever the nature of the initial distribution function of the parameter θ may be.

Conceive of drawing a large number of sets of random samples of N items each from a population consisting either of an infinite supply or of a finite supply with replacements, and that one of these samples, taken at random, yields a value of $t = t_0$ for a certain *statistic*, then the line $t = t_0$ parallel to the θ-axis would fail to intersect the boundaries of the confidence belt, in two points, in at most a small fractional part (less than 2ϵ) of the total number of sets of drawings. Denote the ordinates of the points in which the line $t = t_0$ cuts the curves $C_{1-\epsilon}$ and C_ϵ by θ_1 and θ_2, respectively (Figure 23). These boundary values of θ are called *fiducial limits* of θ that correspond to $t = t_0$ and the interval θ_1 to θ_2 is called the fiducial interval for $t = t_0$. It is important to emphasize that the statement that $1 - 2\epsilon$ is the probability that a value of θ taken at random will fall into the confidence belt is to be associated with the whole belt, that is, with results of repeated application of a sampling procedure to all values of t met with in an extensive statistical experiment, and not merely with an assigned t. The probability that (θ, t) falls within the confidence belt may differ for different assignments of t, but in the long run of statistical experience, the expected relative frequency of points within the confidence belt is $1 - 2\epsilon$. By choos-

ing ϵ to be small, the probability is nearly 1 that the parameter lies within the confidence belt.

The theory of confidence belts and fiducial intervals finds its main application in the testing of a certain hypothesis for possible rejection under the assumption that it is true. Such an hypothesis has been termed a *null hypothesis*. If, for a given ϵ, the null hypothesis is rejected due to the value of t found from the actual data, the value of t is said to be significant at the level of probability equal to 2ϵ. On the other hand, a value of t from observed data which does not reject the null hypothesis is said to be non-significant.

9. Fiducial Limits. (*a*) *For the mean.* Let $\bar{x}$ and s be the mean and standard deviation of a sample of $N = n + 1$ items drawn from a normal universe with unknown mean $\tilde{x}$. The problem is to determine an interval surrounding $\bar{x}$ in which we may assume, with a certain degree of confidence, that $\tilde{x}$ is contained. We learned in Chapter VII that the variable

$$t = \frac{\sqrt{n}(\bar{x} - \tilde{x})}{s} \tag{24}$$

is distributed in accord with the $F_n(t)$ curve and that $P = 1 - P_n(t)$ has been tabulated for various values of t and n, where

$$P_n(t) = 2\int_0^t F_n(t)\,dt.$$

Therefore, for an assigned ϵ and for an assigned value of n, ($n \leq 30$), we may obtain from the tables upper and lower critical values of t by solving the equation $P = 2\epsilon$. With these critical values we can determine from (24) the required interval surrounding $\bar{x}$ for the given value of ϵ. It is conventional among certain workers to take $\epsilon = .005$ (or .025) since they wish to determine values of the estimates of $\tilde{x}$ in an interval dividing hypotheses that will be rejected from those acceptable under a null hypothesis at the 1% (or 5%) level of significance.

Suppose, then, that we make the claim

$$\bar{x} - t_\epsilon \frac{s}{\sqrt{n}} < \tilde{x} < \bar{x} + t_\epsilon \frac{s}{\sqrt{n}} \tag{25}$$

and we desire the probability of an error in this statement to be not more than $2\epsilon = .01$. Taking $n = 15$, for example, we find from

Table 12, Chapter VII, that $t = \pm 2.947$ when $P = .01$. Then we have

$$(\bar{x} - \tilde{x}) = \frac{\pm 2.947s}{\sqrt{15}}$$
$$= \pm .76s$$

and the claim

$$\bar{x} - .76s < \tilde{x} < \bar{x} + .76s$$

will be correct 99% of the time.

It is clear from the above procedure that our confidence in the fiducial limits $\bar{x} \pm t_\epsilon s/\sqrt{n}$ is measured by the area under the $F_n(t)$ curve inside $t = \pm t_\epsilon$, that is, by $P_n(t_\epsilon)$. This means that if we could observe all possible samples, the proportion represented by $P_n(t_\epsilon)$ would yield values of $\bar{x}$ and s for which the claim (25) is true, while the remaining proportion, $P = 1 - P_n(t_\epsilon)$, would yield values of $\bar{x}$ and s for which the claim is false.

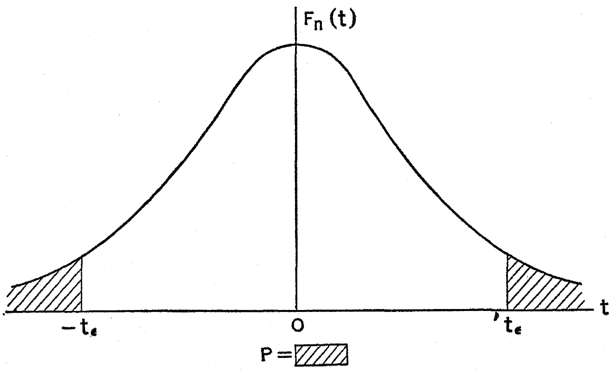


FIG. 24

If we were testing a hypothetical value of $\tilde{x}$ we would say that $\bar{x}$ is not significant at the 1% level of significance if $\tilde{x}$ has any value in the $\bar{x} \pm t_\epsilon s/\sqrt{n}$ interval, $\epsilon = .005$. If $\tilde{x}$ does not lie in this interval we say that $\bar{x}$ is significant at this level.

Obviously, values of t satisfying the equation $P = .01$, that is, $P_n(t) = .99$, vary with n. To avoid the trouble of entering a table we give an *alternate method* which is valid when the sample is not small. Recall that the variable

$$t = \frac{(\bar{x} - \tilde{x})\sqrt{N - 3}}{s}$$

is approximately normally distributed when $N > 30$. The area under the normal curve outside $t = \pm 2.576$ is .01. Therefore, the 99% fiducial range of $\tilde{x}$ is then

$$\bar{x} \pm \frac{2.576s}{\sqrt{N-3}}$$

and the range gets smaller as N increases.

(*b*) *For the difference between two means.* Let $\bar{x}_1$ and $s_1{}^2$ be the observed mean and variance of a sample of N_1 drawn from a normal universe with unknown mean $\tilde{x}_1$ and let $\bar{x}_2$ and $s_2{}^2$ be the observed mean and variance of a sample of N_2 drawn from a normal universe with unknown mean $\tilde{x}_2$. It is assumed that the two universes have a common variance σ^2. For brevity, let

$$\bar{w} = \bar{x}_1 - \bar{x}_2, \qquad \tilde{w} = \tilde{x}_1 - \tilde{x}_2, \qquad N = N_1 + N_2$$

$$\sigma_{\hat{w}} = \left[\left\{\frac{N_1 s_1{}^2 + N_2 s_2{}^2}{N-2}\right\}\left\{\frac{N_1 + N_2}{N_1 N_2}\right\}\right]^{1/2}.$$

Then

$$t = \frac{\bar{w} - \tilde{w}}{\sigma_{\bar{w}}} \tag{26}$$

is distributed in accord with $F_n(t)$ for $n = N - 2$. From (26), upper and lower fiducial values of $\tilde{w}$ can be found by assigning to t the solutions of $P_n(t) = .99$, that is, of $P = .01$. If the value $\tilde{w} = 0$ falls outside the fiducial interval thus established, the conclusion is that the difference between the means is significant at the 1% level. That is, $\tilde{w} \neq 0$ and hence $\tilde{x}_1 \neq \tilde{x}_2$.

If the two samples are equal in number so that the variates can be paired in some manner we may compute (26) by a different method. Let $N = N_1 = N_2$, $w = x_1 - x_2$, and compute $\bar{w}$ and $\sum_1^N (w - \bar{w})^2$.

Then

$$t = \frac{\bar{w} - \tilde{w}}{\sigma_w}$$

$$= \frac{\bar{w} - \tilde{w}}{\dfrac{s_w}{\sqrt{N-1}}}$$

$$= \frac{\bar{w} - \tilde{w}}{\left[\dfrac{\sum_{1}^{N} (w - \bar{w})^2}{N(N-1)}\right]^{1/2}} . \tag{27}$$

The last expression is sometimes called *Bessel's Formula.*

Example 5. (Snedecor[12]) Imagine a newly discovered apple, attractive in appearance, delicious in flavor, having apparently all the qualifications of success. It has been christened " King." Only its yielding capacities in various localities is yet to be tested. The following procedure is decided upon. King is planted adjacent to Standard in 15 orchards scattered about the region suitable for production. Years later, when the trees have matured, the yields are measured and recorded in the following table where x_1 refers to King, x_2 to Standard, and $w = x_1 - x_2$. The yields are in bushels.

x_1	x_2	w	$(w - \bar{w})^2$
13	11	2	16
12	6	6	0
10	3	7	1
6	1	5	1
13	7	6	0
15	10	5	1
19	9	10	16
10	4	6	0
11	3	8	4
11	6	5	1
13	8	5	1
9	5	4	4
14	7	7	1
12	6	6	0
12	4	8	4
	Totals	90	50

Substituting in (27) we get

$$t = \frac{6 - \tilde{w}}{\left[\dfrac{50}{(15)(14)}\right]^{1/2}} = \frac{6 - \tilde{w}}{.488} .$$

Interpolating in Table III for $n = 14$ and checking the result in the more extensive table in Fisher's text we find that $P = .01$ when $t = 2.977$. Then solving the equation

$$\frac{6 - \tilde{w}}{.488} = \pm 2.977$$

we obtain $\tilde{w} = 4.55$ and $\tilde{w} = 7.45$. Since $\tilde{w} = 0$ is outside the interval from 4.55 to 7.45, the observed value of $\bar{w}$ differs significantly from either value of the parameter. In other words, for these as well as for all values outside the fiducial interval 4.55–7.45, we would reject (at the 1% level of significance) the null hypothesis that there is no significant difference between the yields of the two varieties, insofar as their means provide a criterion of judgment.

(c) *For the variance.* In (25) of Chapter VII we obtained the distribution of s^2 which we will now write in the form

$$H(s^2)\,ds^2 = \frac{e^{-Ns^2/2\sigma^2}\left(\dfrac{Ns^2}{2\sigma^2}\right)^{(N-3)/2}}{\left(\dfrac{2\sigma^2}{N}\right)\Gamma\left(\dfrac{N-1}{2}\right)}\,ds^2.$$

If we let $\chi^2 = Ns^2/\sigma^2$ we get the χ^2 distribution given in (12) with N replacing k,

$$T(\chi^2)\,d\chi^2 = \frac{e^{-x^2/2}(\chi^2)^{(N-3)/2}}{2^{(N-1)/2}\Gamma\left(\dfrac{N-1}{2}\right)}\,d\chi^2.$$

That we should thus obtain (12) is more than a coincidence, because it turns out that Ns^2/σ^2 actually is χ^2 for N observations made on a single magnitude. If now we let $n = N - 1$ we obtain the distribution for n degrees of freedom,

$$T_n(\chi^2)\,d\chi^2 = \frac{e^{-x^2/2}(\chi^2)^{(n-2)/2}}{2^{n/2}\Gamma\left(\dfrac{n}{2}\right)}\,d\chi^2. \tag{28}$$

To determine the fiducial limits of σ^2 we first observe from (3) of Chapter VII that $Ns^2 = n\hat{\sigma}^2 = \sum_1^N (x_i - \bar{x})^2$, and therefore we may write $\chi^2 = n\hat{\sigma}^2/\sigma^2$. If now we make the claim

$$\frac{n\hat{\sigma}^2}{\chi_2^2} < \sigma^2 < \frac{n\hat{\sigma}^2}{\chi_1^2},$$

where χ_1^2 and χ_2^2 are arbitrarily chosen constants ($\chi_1^2 < \chi_2^2$), then our " measure of confidence " in the correctness of this claim is given by $I_n(\chi_1^2) - I_n(\chi_2^2)$, where

$$I_n(\chi^2) = \int_{\chi^2}^{\infty} T_n(\chi^2)\,d\chi^2.$$

Values of $I_n(\chi^2)$ can be obtained from Pearson's *Tables.*[3]

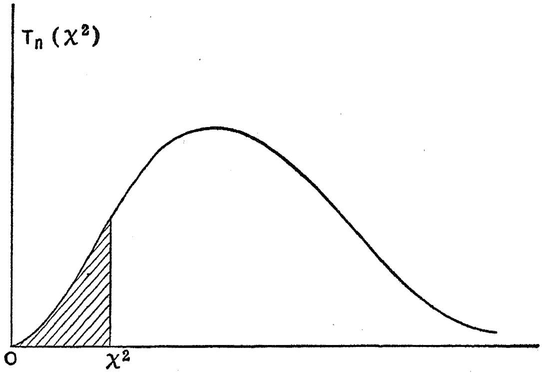


FIG. 25

For further study of fiducial inference and its applications to testing hypotheses, the reader is referred to the publications of Fisher,[9] Neyman,[10] and Wilks.[11]

Notes and References

1. In preparing §§ 1 and 2, the writer has derived much help from *Probability and Its Engineering Uses*, Fry, D. Van Nostrand Co., and *Statistical Inference, 1936–37*, Wilks, Edwards Brothers.
2. This part of § 1 and practically all of § 2 are taken from Professor Wilks' *Lectures* and reproduced with his permission.
3. *Tables for Statisticians and Biometricians.*
4. R. A. Fisher, *Statistical Methods for Research Workers.*
5. Karl Pearson, *On the Criterion that a Given Set of Deviations from the Probable in the Case of Correlated Variables is Such that Can Reasonably be Supposed to have Arisen from Random Sampling*, Phil. Mag. 5th Series, vol. 50 (1900), pp. 157–175.
6. R. A. Fisher, *Uncertain Inference*, Proc. Amer. Acad. Arts and Sciences, vol. 71, no. 4, pp. 245–257.
7. E. C. Molina, *Bayes Theorem*, Annals of Mathematical Statistics, vol. 2, no. 1, pp. 23–37.
8. H. L. Rietz, *On a Certain Advance in Statistical Inference*, American Mathematical Monthly, vol. 45, pp. 149–158.
9. R. A. Fisher, *Inverse Probability*, Proc. Camb. Phil. Soc., vol. 26, 1930, p. 528; Proc. Royal Soc., A, vol. 139, 1933, p. 343.
10. J. Neyman, *On Different Aspects of the Representative Method, the Method of Stratified Sampling, and the Method of Purposive Selection*, Journal Royal Statistical Society, vol. 97, 1934, pp. 558–606.
11. S. S. Wilks, (*a*) *Lectures on Statistical Inference, 1936–1937.*
    (*b*) *Fiducial Distributions in Fiducial Inference*, Annals of Mathematical Statistics, vol. IX, no. 4, pp. 272–280. (This is an expository paper.)
12. G. W. Snedecor, *Statistical Methods.* Collegiate Press, Inc., Ames, Iowa.

Exercises

1. Read the following paper: *The* χ^2*-Test of Significance,* T. C. Fry, Journal of the American Statistical Association, vol. 33, pp. 513–525. (The three papers following Fry's exposition are also recommended.)
2. Toss seven coins 128 times and record the frequencies of heads. Apply the χ^2-test to the resulting distribution.
3. Graduate an appropriate distribution in Part I by means of the normal curve and test the composite hypothesis that the observed distribution was a sample from a normal universe having the mean and standard deviation of the sample.
4. Give a report on χ^2 and contingency tables.
5. (*Chrystal*) A bag contains three balls, each of which is either white or black, all possible numbers of white being equally likely. Two at once are drawn at random and prove to be white. What is the probability that all of the balls are white? *Ans.* $\frac{3}{4}$.
6. If, in Example 4, it is assumed that, initially, all possible numbers of white balls in the urn are equally likely, what is the solution?
7. If N is large show that the 95% fiducial range of $\bar{x}$ for a normal universe is $\bar{x} \pm 1.96/\sqrt{N-3}$.
8. Making use of the references cited prepare a report on fiducial inference.

Review Problem

A question arose in a physical education class as to whether eleven-year-old girls weigh, as a rule, more than eleven-year-old boys. Suppose you wished to make a thorough analysis of the data in the table below concerning weights of boys and girls aged eleven. Describe the tests you might apply, the reasoning and assumptions underlying these, and the interpretation that might be placed on the results.

Weight (pounds)	*Frequency*	
Class Marks	*Boys*	*Girls*
42.5	1	0
48.5	3	1
54.5	9	7
60.5	33	37
66.5	65	41
72.5	80	59
78.5	72	58
84.5	41	48
90.5	27	23
96.5	7	26
102.5	4	16
108.5	2	5
114.5	1	3
120.5	0	2
Totals	345	326

The following points are suggested for discussion:

(*a*) Is there a clear difference between the two distributions? How would you test this: from the means, from the variances, from the samples as a whole?

(*b*) 32.3% of the boys and 26.4% of the girls have weights less than 69.5 pounds. Is this difference significant?

(*c*) Within what limits would you say that the mean and standard deviation in the population of eleven-year-old boys (from which you have the sample of 345) is almost certain to lie in each case?

(*d*) Summarize your results.

APPENDIX

Tables

I. Ordinates and Areas of the Normal Curve.

II. 5% and 1% Points for the Distribution of F.

III. χ^2 Probability Scale.

Table I. Ordinates and Areas of the Normal Curve, $\phi(t) = \dfrac{1}{\sqrt{2\pi}} e^{-t^2/2}$

t	$\phi(t)$	$\int_0^t \phi(t)dt$	t	$\phi(t)$	$\int_0^t \phi(t)dt$	t	$\phi(t)$	$\int_0^t \phi(t)dt$
.00	.39894	.00000	.45	.36053	.17364	.90	.26609	.31594
.01	.39892	.00399	.46	.35889	.17724	.91	.26369	.31859
.02	.39886	.00798	.47	.35723	.18082	.92	.26129	.32121
.03	.39876	.01197	.48	.35553	.18439	.93	.25888	.32381
.04	.39862	.01595	.49	.35381	.18793	.94	.25647	.32639
.05	.39844	.01994	.50	.35207	.19146	.95	.25406	.32894
.06	.39822	.02392	.51	.35029	.19497	.96	.25164	.33147
.07	.39797	.02790	.52	.34849	.19847	.97	.24923	.33398
.08	.39767	.03188	.53	.34667	.20194	.98	.24681	.33646
.09	.39733	.03586	.54	.34482	.20540	.99	.24439	.33891
.10	.39695	.03983	.55	.34294	.20884	1.00	.24197	.34134
.11	.39654	.04380	.56	.34105	.21226	1.01	.23955	.34375
.12	.39608	.04776	.57	.33912	.21566	1.02	.23713	.34614
.13	.39559	.05172	.58	.33718	.21904	1.03	.23471	.34850
.14	.39505	.05567	.59	.33521	.22240	1.04	.23230	.35083
.15	.39448	.05962	.60	.33322	.22575	1.05	.22988	.35314
.16	.39387	.06356	.61	.33121	.22907	1.06	.22747	.35543
.17	.39322	.06749	.62	.32918	.23237	1.07	.22506	.35769
.18	.39253	.07142	.63	.32713	.23565	1.08	.22265	.35993
.19	.39181	.07535	.64	.32506	.23891	1.09	.22025	.36214
.20	.39104	.07926	.65	.32297	.24215	1.10	.21785	.36433
.21	.39024	.08317	.66	.32086	.24537	1.11	.21546	.36650
.22	.38940	.08706	.67	.31874	.24857	1.12	.21307	.36864
.23	.38853	.09095	.68	.31659	.25175	1.13	.21069	.37076
.24	.38762	.09483	.69	.31443	.25490	1.14	.20831	.37286
.25	.38667	.09871	.70	.31225	.25804	1.15	.20594	.37493
.26	.38568	.10257	.71	.31006	.26115	1.16	.20357	.37698
.27	.38466	.10642	.72	.30785	.26424	1.17	.20121	.37900
.28	.38361	.11026	.73	.30563	.26730	1.18	.19886	.38100
.29	.38251	.11409	.74	.30339	.27035	1.19	.19652	.38298
.30	.38139	.11791	.75	.30114	.27337	1.20	.19419	.38493
.31	.38023	.12172	.76	.29887	.27637	1.21	.19186	.38686
.32	.37903	.12552	.77	.29659	.27935	1.22	.18954	.38877
.33	.37780	.12930	.78	.29431	.28230	1.23	.18724	.39065
.34	.37654	.13307	.79	.29200	.28524	1.24	.18494	.39251
.35	.37524	.13683	.80	.28969	.28814	1.25	.18265	.39435
.36	.37391	.14058	.81	.28737	.29103	1.26	.18037	.39617
.37	.37255	.14431	.82	.28504	.29389	1.27	.17810	.39796
.38	.37115	.14803	.83	.28269	.29673	1.28	.17585	.39973
.39	.36973	.15173	.84	.28034	.29955	1.29	.17360	.40147
.40	.36827	.15542	.85	.27798	.30234	1.30	.17137	.40320
.41	.36678	.15910	.86	.27562	.30511	1.31	.16915	.40490
.42	.36526	.16276	.87	.27324	.30785	1.32	.16694	.40658
.43	.36371	.16640	.88	.27086	.31057	1.33	.16474	.40824
.44	.36213	.17003	.89	.26848	.31327	1.34	.16256	.40988

Table I. Ordinates and Areas of the Normal Curve, $\phi(t) = \dfrac{1}{\sqrt{2\pi}} e^{-t^2/2}$

t	$\phi(t)$	$\int_0^t \phi(t)\,dt$	t	$\phi(t)$	$\int_0^t \phi(t)\,dt$	t	$\phi(t)$	$\int_0^t \phi(t)\,dt$
1.35	.16038	.41149	1.80	.07895	.46407	2.25	.03174	.48778
1.36	.15822	.41309	1.81	.07754	.46485	2.26	.03103	.48809
1.37	.15608	.41466	1.82	.07614	.46562	2.27	.03034	.48840
1.38	.15395	.41621	1.83	.07477	.46638	2.28	.02965	.48870
1.39	.15183	.41774	1.84	.07341	.46712	2.29	.02898	.48899
1.40	.14973	.41924	1.85	.07206	.46784	2.30	.02833	.48928
1.41	.14764	.42073	1.86	.07074	.46856	2.31	.02768	.48956
1.42	.14556	.42220	1.87	.06943	.46926	2.32	.02705	.48983
1.43	.14350	.42364	1.88	.06814	.46995	2.33	.02643	.49010
1.44	.14146	.42507	1.89	.06687	.47062	2.34	.02582	.49036
1.45	.13943	.42647	1.90	.06562	.47128	2.35	.02522	.49061
1.46	.13742	.42786	1.91	.06439	.47193	2.36	.02463	.49086
1.47	.13542	.42922	1.92	.06316	.47257	2.37	.02406	.49111
1.48	.13344	.43056	1.93	.06195	.47320	2.38	.02349	.49134
1.49	.13147	.43189	1.94	.06077	.47381	2.39	.02294	.49158
1.50	.12952	.43319	1.95	.05959	.47441	2.40	.02239	.49180
1.51	.12758	.43448	1.96	.05844	.47500	2.41	.02186	.49202
1.52	.12566	.43574	1.97	.05730	.47558	2.42	.02134	.49224
1.53	.12376	.43699	1.98	.05618	.47615	2.43	.02083	.49245
1.54	.12188	.43822	1.99	.05508	.47670	2.44	.02033	.49266
1.55	.12001	.43943	2.00	.05399	.47725	2.45	.01984	.49286
1.56	.11816	.44062	2.01	.02592	.47778	2.46	.01936	.49305
1.57	.11632	.44179	2.02	.05186	.47831	2.47	.01889	.49324
1.58	.11450	.44295	2.03	.05082	.47882	2.48	.01842	.49343
1.59	.11270	.44408	2.04	.04980	.47932	2.49	.01797	.49361
1.60	.11092	.44520	2.05	.04879	.47982	2.50	.01753	.49379
1.61	.10915	.44630	2.06	.04780	.48030	2.51	.01709	.49396
1.62	.10741	.44738	2.07	.04682	.48077	2.52	.01667	.49413
1.63	.10567	.44845	2.08	.04586	.48124	2.53	.01625	.49430
1.64	.10396	.44950	2.09	.04491	.48169	2.54	.01585	.49446
1.65	.10226	.45053	2.10	.04398	.48214	2.55	.01545	.49461
1.66	.10059	.45154	2.11	.04307	.48257	2.56	.01506	.49477
1.67	.09893	.45254	2.12	.04217	.48300	2.57	.01468	.49492
1.68	.09728	.45352	2.13	.04128	.48341	2.58	.01431	.49506
1.69	.09566	.45449	2.14	.04041	.48382	2.59	.01394	.49520
1.70	.09405	.45543	2.15	.03955	.48422	2.60	.01358	.49534
1.71	.09246	.45637	2.16	.03871	.48461	2.61	.01323	.49547
1.72	.09089	.45728	2.17	.03788	.48500	2.62	.01289	.49560
1.73	.08933	.45818	2.18	.03706	.48537	2.63	.01256	.49573
1.74	.08780	.45907	2.19	.03626	.48574	2.64	.01223	.49585
1.75	.08628	.45994	2.20	.03547	.48610	2.65	.01191	.49598
1.76	.08478	.46080	2.21	.03470	.48645	2.66	.01160	.49609
1.77	.08329	.46164	2.22	.03394	.48679	2.67	.01130	.49621
1.78	.08183	.46246	2.23	.03319	.48713	2.68	.01100	.49632
1.79	.08038	.46327	2.24	.03246	.48745	2.69	.01071	.49643

TABLE I. ORDINATES AND AREAS OF THE NORMAL CURVE, $\phi(t) = \dfrac{1}{\sqrt{2\pi}} e^{-t^2/2}$

t	$\phi(t)$	$\int_0^t \phi(t)dt$	t	$\phi(t)$	$\int_0^t \phi(t)dt$	t	$\phi(t)$	$\int_0^t \phi(t)dt$
2.70	.01042	.49653	3.15	.00279	.49918	3.60	.00061	.49984
2.71	.01014	.49664	3.16	.00271	.49921	3.61	.00059	.49985
2.72	.00987	.49674	3.17	.00262	.49924	3.62	.00057	.49985
2.73	.00961	.49683	3.18	.00254	.49926	3.63	.00055	.49986
2.74	.00935	.49693	3.19	.00246	.49929	3.64	.00053	.49986
2.75	.00909	.49702	3.20	.00238	.49391	3.65	.00051	.49987
2.76	.00885	.49711	3.21	.00231	.49934	3.66	.00049	.49987
2.77	.00861	.49720	3.22	.00224	.49936	3.67	.00047	.49988
2.78	.00837	.49728	3.23	.00216	.49938	3.68	.00046	.49988
2.79	.00814	.49736	3.24	.00210	.49940	3.69	.00044	.49989
2.80	.00792	.49744	3.25	.00203	.49942	3.70	.00042	.49989
2.81	.00770	.49752	3.26	.00196	.49944	3.71	.00041	.49990
2.82	.00748	.49760	3.27	.00190	.49946	3.72	.00039	.49990
2.83	.00727	.49767	3.28	.00184	.49948	3.73	.00038	.49990
2.84	.00707	.49774	3.29	.00178	.49950	3.74	.00037	.49991
2.85	.00687	.49781	3.30	.00172	.49952	3.75	.00035	.49991
2.86	.00668	.49788	3.31	.00167	.49953	3.76	.00034	.49992
2.87	.00649	.49795	3.32	.00161	.49955	3.77	.00033	.49992
2.88	.00631	.49801	3.33	.00156	.49957	3.78	.00031	.49992
2.89	.00613	.49807	3.34	.00151	.49958	3.79	.00030	.49992
2.90	.00595	.49813	3.35	.00146	.49960	3.80	.00029	.49993
2.91	.00578	.49819	3.36	.00141	.49961	3.81	.00028	.49993
2.92	.00562	.49825	3.37	.00136	.49962	3.82	.00027	.49993
2.93	.00545	.49831	3.38	.00132	.49964	3.83	.00026	.49994
2.94	.00530	.49836	3.39	.00127	.49965	3.84	.00025	.49994
2.95	.00514	.49841	3.40	.00123	.49966	3.85	.00024	.49994
2.96	.00499	.49846	3.41	.00119	.49968	3.86	.00023	.49994
2.97	.00485	.49851	3.42	.00115	.49969	3.87	.00022	.49995
2.98	.00471	.49856	3.43	.00111	.49970	3.88	.00021	.49995
2.99	.00457	.49861	3.44	.00107	.49971	3.89	.00021	.49995
3.00	.00443	.49865	3.45	.00104	.49972	3.90	.00020	.49995
3.01	.00430	.49869	3.46	.00100	.49973	3.91	.00019	.49995
3.02	.00417	.49874	3.47	.00097	.49974	3.92	.00018	.49996
3.03	.00405	.49878	3.48	.00094	.49975	3.93	.00018	.49996
3.04	.00393	.49882	3.49	.00090	.49976	3.94	.00017	.49996
3.05	.00381	.49886	3.50	.00087	.49977	3.95	.00016	.49996
3.06	.00370	.49889	3.51	.00084	.49978	3.96	.00016	.49996
3.07	.00358	.49893	3.52	.00081	.49978	3.97	.00015	.49996
3.08	.00348	.49897	3.53	.00079	.49979	3.98	.00014	.49997
3.09	.00337	.49900	3.54	.00076	.49980	3.99	.00014	.49997
3.10	.00327	.49903	3.55	.00073	.49981			
3.11	.00317	.49906	3.56	.00071	.49981			
3.12	.00307	.49910	3.57	.00068	.49982			
3.13	.00298	.49913	3.58	.00066	.49983			
3.14	.00288	.49916	3.59	.00063	.49983			

Table II.* 5% (Roman Type) and 1% (Bold Face Type) Points for the Distribution of F

n_2	n_1 degrees of freedom (for greater mean square)																								n_2
	1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	24	30	40	50	75	100	200	500	∞	
1	161 **4,052**	200 **4,999**	216 **5,403**	225 **5,625**	230 **5,764**	234 **5,859**	237 **5,928**	239 **5,981**	241 **6,022**	242 **6,056**	243 **6,082**	244 **6,106**	245 **6,142**	246 **6,169**	248 **6,208**	249 **6,234**	250 **6,258**	251 **6,286**	252 **6,302**	253 **6,323**	253 **6,334**	254 **6,352**	254 **6,361**	254 **6,366**	1
2	18.51 **98.49**	19.00 **99.01**	19.16 **99.17**	19.25 **99.25**	19.30 **99.30**	19.33 **99.33**	19.36 **99.34**	19.37 **99.36**	19.38 **99.38**	19.39 **99.40**	19.40 **99.41**	19.41 **99.42**	19.42 **99.43**	19.43 **99.44**	19.44 **99.45**	19.45 **99.46**	19.46 **99.47**	19.47 **99.48**	19.47 **99.48**	19.48 **99.49**	19.49 **99.49**	19.49 **99.49**	19.50 **99.50**	19.50 **99.50**	2
3	10.13 **34.12**	9.55 **30.81**	9.28 **29.46**	9.12 **28.71**	9.01 **28.24**	8.94 **27.91**	8.88 **27.67**	8.84 **27.49**	8.81 **27.34**	8.78 **27.23**	8.76 **27.13**	8.74 **27.05**	8.71 **26.92**	8.69 **26.83**	8.66 **26.69**	8.64 **26.60**	8.62 **26.50**	8.60 **26.41**	8.58 **26.35**	8.57 **26.27**	8.56 **26.23**	8.54 **26.18**	8.54 **26.14**	8.53 **26.12**	3
4	7.71 **21.20**	6.94 **18.00**	6.59 **16.69**	6.39 **15.98**	6.26 **15.52**	6.16 **15.21**	6.09 **14.98**	6.04 **14.80**	6.00 **14.66**	5.96 **14.54**	5.93 **14.45**	5.91 **14.37**	5.87 **14.24**	5.84 **14.15**	5.80 **14.02**	5.77 **13.93**	5.74 **13.83**	5.71 **13.74**	5.70 **13.69**	5.68 **13.61**	5.66 **13.57**	5.65 **13.52**	5.64 **13.48**	5.63 **13.46**	4
5	6.61 **16.26**	5.79 **13.27**	5.41 **12.06**	5.19 **11.39**	5.05 **10.97**	4.95 **10.67**	4.88 **10.45**	4.82 **10.27**	4.78 **10.15**	4.74 **10.05**	4.70 **9.96**	4.68 **9.89**	4.64 **9.77**	4.60 **9.68**	4.56 **9.55**	4.53 **9.47**	4.50 **9.38**	4.46 **9.29**	4.44 **9.24**	4.42 **9.17**	4.40 **9.13**	4.38 **9.07**	4.37 **9.04**	4.36 **9.02**	5
6	5.99 **13.74**	5.14 **10.92**	4.76 **9.78**	4.53 **9.15**	4.39 **8.75**	4.28 **8.47**	4.21 **8.26**	4.15 **8.10**	4.10 **7.98**	4.06 **7.87**	4.03 **7.79**	4.00 **7.72**	3.96 **7.60**	3.92 **7.52**	3.87 **7.39**	3.84 **7.31**	3.81 **7.23**	3.77 **7.14**	3.75 **7.09**	3.72 **7.02**	3.71 **6.99**	3.69 **6.94**	3.68 **6.90**	3.67 **6.88**	6
7	5.59 **12.25**	4.74 **9.55**	4.35 **8.45**	4.12 **7.85**	3.97 **7.46**	3.87 **7.19**	3.79 **7.00**	3.73 **6.84**	3.68 **6.71**	3.63 **6.62**	3.60 **6.54**	3.57 **6.47**	3.52 **6.35**	3.49 **6.27**	3.44 **6.15**	3.41 **6.07**	3.38 **5.98**	3.34 **5.90**	3.32 **5.85**	3.29 **5.78**	3.28 **5.75**	3.25 **5.70**	3.24 **5.67**	3.23 **5.65**	7
8	5.32 **11.26**	4.46 **8.65**	4.07 **7.59**	3.84 **7.01**	3.69 **6.63**	3.58 **6.37**	3.50 **6.19**	3.44 **6.03**	3.39 **5.91**	3.34 **5.82**	3.31 **5.74**	3.28 **5.67**	3.23 **5.56**	3.20 **5.48**	3.15 **5.36**	3.12 **5.28**	3.08 **5.20**	3.05 **5.11**	3.03 **5.06**	3.00 **5.00**	2.98 **4.96**	2.96 **4.91**	2.94 **4.88**	2.93 **4.86**	8
9	5.12 **10.56**	4.26 **8.02**	3.86 **6.99**	3.63 **6.42**	3.48 **6.06**	3.37 **5.80**	3.29 **5.62**	3.23 **5.47**	3.18 **5.35**	3.13 **5.26**	3.10 **5.18**	3.07 **5.11**	3.02 **5.00**	2.98 **4.92**	2.93 **4.80**	2.90 **4.73**	2.86 **4.64**	2.82 **4.56**	2.80 **4.51**	2.77 **4.45**	2.76 **4.41**	2.73 **4.36**	2.72 **4.33**	2.71 **4.31**	9
10	4.96 **10.04**	4.10 **7.56**	3.71 **6.55**	3.48 **5.99**	3.33 **5.64**	3.22 **5.39**	3.14 **5.21**	3.07 **5.06**	3.02 **4.95**	2.97 **4.85**	2.94 **4.78**	2.91 **4.71**	2.86 **4.60**	2.82 **4.52**	2.77 **4.41**	2.74 **4.33**	2.70 **4.25**	2.67 **4.17**	2.64 **4.12**	2.61 **4.05**	2.59 **4.01**	2.56 **3.96**	2.55 **3.93**	2.54 **3.91**	10
11	4.84 **9.65**	3.98 **7.20**	3.59 **6.22**	3.36 **5.67**	3.20 **5.32**	3.09 **5.07**	3.01 **4.88**	2.95 **4.74**	2.90 **4.63**	2.86 **4.54**	2.82 **4.46**	2.79 **4.40**	2.74 **4.29**	2.70 **4.21**	2.65 **4.10**	2.61 **4.02**	2.57 **3.94**	2.53 **3.86**	2.50 **3.80**	2.47 **3.74**	2.45 **3.70**	2.42 **3.66**	2.41 **3.62**	2.40 **3.60**	11
12	4.75 **9.33**	3.88 **6.93**	3.49 **5.95**	3.26 **5.41**	3.11 **5.06**	3.00 **4.82**	2.92 **4.65**	2.85 **4.50**	2.80 **4.39**	2.76 **4.30**	2.72 **4.22**	2.69 **4.16**	2.64 **4.05**	2.60 **3.98**	2.54 **3.86**	2.50 **3.78**	2.46 **3.70**	2.42 **3.61**	2.40 **3.56**	2.36 **3.49**	2.35 **3.46**	2.32 **3.41**	2.31 **3.38**	2.30 **3.36**	12
13	4.67 **9.07**	3.80 **6.70**	3.41 **5.74**	3.18 **5.20**	3.02 **4.86**	2.92 **4.62**	2.84 **4.44**	2.77 **4.30**	2.72 **4.19**	2.67 **4.10**	2.63 **4.02**	2.60 **3.96**	2.55 **3.85**	2.51 **3.78**	2.46 **3.67**	2.42 **3.59**	2.38 **3.51**	2.34 **3.42**	2.32 **3.37**	2.28 **3.30**	2.26 **3.27**	2.24 **3.21**	2.22 **3.18**	2.21 **3.16**	13

* Reproduced from *Statistical Methods* by G. W. Snedecor by permission of the author and the publisher, Collegiate Press, Inc., Ames, Iowa.

TABLE II. 5% (ROMAN TYPE) AND 1% (BOLD FACE TYPE) POINTS FOR THE DISTRIBUTION OF F

n_2	n_1 degrees of freedom (for greater mean square)																								n_2
	1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	24	30	40	50	75	100	200	500	∞	
14	4.60 **8.86**	3.74 **6.51**	3.34 **5.56**	3.11 **5.03**	2.96 **4.69**	2.85 **4.46**	2.77 **4.28**	2.70 **4.14**	2.65 **4.03**	2.60 **3.94**	2.56 **3.86**	2.53 **3.80**	2.48 **3.70**	2.44 **3.62**	2.39 **3.51**	2.35 **3.43**	2.31 **3.34**	2.27 **3.26**	2.24 **3.21**	2.21 **3.14**	2.19 **3.11**	2.16 **3.06**	2.14 **3.02**	2.13 **3.00**	14
15	4.54 **8.68**	3.68 **6.36**	3.29 **5.42**	3.06 **4.89**	2.90 **4.56**	2.79 **4.32**	2.70 **4.14**	2.64 **4.00**	2.59 **3.89**	2.55 **3.80**	2.51 **3.73**	2.48 **3.67**	2.43 **3.56**	2.39 **3.48**	2.33 **3.36**	2.29 **3.29**	2.25 **3.20**	2.21 **3.12**	2.18 **3.07**	2.15 **3.00**	2.12 **2.97**	2.10 **2.92**	2.08 **2.89**	2.07 **2.87**	15
16	4.49 **8.53**	3.63 **6.23**	3.24 **5.29**	3.01 **4.77**	2.85 **4.44**	2.74 **4.20**	2.66 **4.03**	2.59 **3.89**	2.54 **3.78**	2.49 **3.69**	2.45 **3.61**	2.42 **3.55**	2.37 **3.45**	2.33 **3.37**	2.28 **3.25**	2.24 **3.18**	2.20 **3.10**	2.16 **3.01**	2.13 **2.96**	2.09 **2.89**	2.07 **2.86**	2.04 **2.80**	2.02 **2.77**	2.01 **2.75**	16
17	4.45 **8.40**	3.59 **6.11**	3.20 **5.18**	2.96 **4.67**	2.81 **4.34**	2.70 **4.10**	2.62 **3.93**	2.55 **3.79**	2.50 **3.68**	2.45 **3.59**	2.41 **3.52**	2.38 **3.45**	2.33 **3.35**	2.29 **3.27**	2.23 **3.16**	2.19 **3.08**	2.15 **3.00**	2.11 **2.92**	2.08 **2.86**	2.04 **2.79**	2.02 **2.76**	1.99 **2.70**	1.97 **2.67**	1.96 **2.65**	17
18	4.41 **8.28**	3.55 **6.01**	3.16 **5.09**	2.93 **4.58**	2.77 **4.25**	2.66 **4.01**	2.58 **3.85**	2.51 **3.71**	2.46 **3.60**	2.41 **3.51**	2.37 **3.44**	2.34 **3.37**	2.29 **3.27**	2.25 **3.19**	2.19 **3.07**	2.15 **3.00**	2.11 **2.91**	2.07 **2.83**	2.04 **2.78**	2.00 **2.71**	1.98 **2.68**	1.95 **2.62**	1.93 **2.59**	1.92 **2.57**	18
19	4.38 **8.18**	3.52 **5.93**	3.13 **5.01**	2.90 **4.50**	2.74 **4.17**	2.63 **3.94**	2.55 **3.77**	2.48 **3.63**	2.43 **3.52**	2.38 **3.43**	2.34 **3.36**	2.31 **3.30**	2.26 **3.19**	2.21 **3.12**	2.15 **3.00**	2.11 **2.92**	2.07 **2.84**	2.02 **2.76**	2.00 **2.70**	1.96 **2.63**	1.94 **2.60**	1.91 **2.54**	1.90 **2.51**	1.88 **2.49**	19
20	4.35 **8.10**	3.49 **5.85**	3.10 **4.94**	2.87 **4.43**	2.71 **4.10**	2.60 **3.87**	2.52 **3.71**	2.45 **3.56**	2.40 **3.45**	2.35 **3.37**	2.31 **3.30**	2.28 **3.23**	2.23 **3.13**	2.18 **3.05**	2.12 **2.94**	2.08 **2.86**	2.04 **2.77**	1.99 **2.69**	1.96 **2.63**	1.92 **2.56**	1.90 **2.53**	1.87 **2.47**	1.85 **2.44**	1.84 **2.42**	20
21	4.32 **8.02**	3.47 **5.78**	3.07 **4.87**	2.84 **4.37**	2.68 **4.04**	2.57 **3.81**	2.49 **3.65**	2.42 **3.51**	2.37 **3.40**	2.32 **3.31**	2.28 **3.24**	2.25 **3.17**	2.20 **3.07**	2.15 **2.99**	2.09 **2.88**	2.05 **2.80**	2.00 **2.72**	1.96 **2.63**	1.93 **2.58**	1.89 **2.51**	1.87 **2.47**	1.84 **2.42**	1.82 **2.38**	1.81 **2.36**	21
22	4.30 **7.94**	3.44 **5.72**	3.05 **4.82**	2.82 **4.31**	2.66 **3.99**	2.55 **3.76**	2.47 **3.59**	2.40 **3.45**	2.35 **3.35**	2.30 **3.26**	2.26 **3.18**	2.23 **3.12**	2.18 **3.02**	2.13 **2.94**	2.07 **2.83**	2.03 **2.75**	1.98 **2.67**	1.93 **2.58**	1.91 **2.53**	1.87 **2.46**	1.84 **2.42**	1.81 **2.37**	1.80 **2.33**	1.78 **2.31**	22
23	4.28 **7.88**	3.42 **5.66**	3.03 **4.76**	2.80 **4.26**	2.64 **3.94**	2.53 **3.71**	2.45 **3.54**	2.38 **3.41**	2.32 **3.30**	2.28 **3.21**	2.24 **3.14**	2.20 **3.07**	2.14 **2.97**	2.10 **2.89**	2.04 **2.78**	2.00 **2.70**	1.96 **2.62**	1.91 **2.53**	1.88 **2.48**	1.84 **2.41**	1.82 **2.37**	1.79 **2.32**	1.77 **2.28**	1.76 **2.26**	23
24	4.26 **7.82**	3.40 **5.61**	3.01 **4.72**	2.78 **4.22**	2.62 **3.90**	2.51 **3.67**	2.43 **3.50**	2.36 **3.36**	2.30 **3.25**	2.26 **3.17**	2.22 **3.09**	2.18 **3.03**	2.13 **2.93**	2.09 **2.85**	2.02 **2.74**	1.98 **2.66**	1.94 **2.58**	1.89 **2.49**	1.86 **2.44**	1.82 **2.36**	1.80 **2.33**	1.76 **2.27**	1.74 **2.23**	1.73 **2.21**	24
25	4.24 **7.77**	3.38 **5.57**	2.99 **4.68**	2.76 **4.18**	2.60 **3.86**	2.49 **3.63**	2.41 **3.46**	2.34 **3.32**	2.28 **3.21**	2.24 **3.13**	2.20 **3.05**	2.16 **2.99**	2.11 **2.89**	2.06 **2.81**	2.00 **2.70**	1.96 **2.62**	1.92 **2.54**	1.87 **2.45**	1.84 **2.40**	1.80 **2.32**	1.77 **2.29**	1.74 **2.23**	1.72 **2.19**	1.71 **2.17**	25
26	4.22 **7.72**	3.37 **5.53**	2.98 **4.64**	2.74 **4.14**	2.59 **3.82**	2.47 **3.59**	2.39 **3.42**	2.32 **3.29**	2.27 **3.17**	2.22 **3.09**	2.18 **3.02**	2.15 **2.96**	2.10 **2.86**	2.05 **2.77**	1.99 **2.66**	1.95 **2.58**	1.90 **2.50**	1.85 **2.41**	1.82 **2.36**	1.78 **2.28**	1.76 **2.25**	1.72 **2.19**	1.70 **2.15**	1.69 **2.13**	26

TABLE II. 5% (ROMAN TYPE) AND 1% (BOLD FACE TYPE) POINTS FOR THE DISTRIBUTION OF F

n_2	n_1 degrees of freedom (for greater mean square)																								n_2
	1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	24	30	40	50	75	100	200	500	∞	
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.30	2.25	2.20	2.16	2.13	2.08	2.03	1.97	1.93	1.88	1.84	1.80	1.76	1.74	1.71	1.68	1.67	27
	7.68	**5.49**	**4.60**	**4.11**	**3.79**	**3.56**	**3.39**	**3.26**	**3.14**	**3.06**	**2.98**	**2.93**	**2.83**	**2.74**	**2.63**	**2.55**	**2.47**	**2.38**	**2.33**	**2.25**	**2.21**	**2.16**	**2.12**	**2.10**	
28	4.20	3.34	2.95	2.71	2.56	2.44	2.36	2.29	2.24	2.19	2.15	2.12	2.06	2.02	1.96	1.91	1.87	1.81	1.78	1.75	1.72	1.69	1.67	1.65	28
	7.64	**5.45**	**4.57**	**4.07**	**3.76**	**3.53**	**3.36**	**3.23**	**3.11**	**3.03**	**2.95**	**2.90**	**2.80**	**2.71**	**2.60**	**2.52**	**2.44**	**2.35**	**2.30**	**2.22**	**2.18**	**2.13**	**2.09**	**2.06**	
29	4.18	3.33	2.93	2.70	2.54	2.43	2.35	2.28	2.22	2.18	2.14	2.10	2.05	2.00	1.94	1.90	1.85	1.80	1.77	1.73	1.71	1.68	1.65	1.64	29
	7.60	**5.42**	**4.54**	**4.04**	**3.73**	**3.50**	**3.33**	**3.20**	**3.08**	**3.00**	**2.92**	**2.87**	**2.77**	**2.68**	**2.57**	**2.49**	**2.41**	**2.32**	**2.27**	**2.19**	**2.15**	**2.10**	**2.06**	**2.03**	
30	4.17	3.32	2.92	2.69	2.53	2.42	2.34	2.27	2.21	2.16	2.12	2.09	2.04	1.99	1.93	1.89	1.84	1.79	1.76	1.72	1.69	1.66	1.64	1.62	30
	7.56	**5.39**	**4.51**	**4.02**	**3.70**	**3.47**	**3.30**	**3.17**	**3.06**	**2.98**	**2.90**	**2.84**	**2.74**	**2.66**	**2.55**	**2.47**	**2.38**	**2.29**	**2.24**	**2.16**	**2.13**	**2.07**	**2.03**	**2.01**	
32	4.15	3.30	2.90	2.67	2.51	2.40	2.32	2.25	2.19	2.14	2.10	2.07	2.02	1.97	1.91	1.86	1.82	1.76	1.74	1.69	1.67	1.64	1.61	1.59	32
	7.50	**5.34**	**4.46**	**3.97**	**3.66**	**3.42**	**3.25**	**3.12**	**3.01**	**2.94**	**2.86**	**2.80**	**2.70**	**2.62**	**2.51**	**2.42**	**2.34**	**2.25**	**2.20**	**2.12**	**2.08**	**2.02**	**1.98**	**1.96**	
34	4.13	3.28	2.88	2.65	2.49	2.38	2.30	2.23	2.17	2.12	2.08	2.05	2.00	1.95	1.89	1.84	1.80	1.74	1.71	1.67	1.64	1.61	1.59	1.57	34
	7.44	**5.29**	**4.42**	**3.93**	**3.61**	**3.38**	**3.21**	**3.08**	**2.97**	**2.89**	**2.82**	**2.76**	**2.66**	**2.58**	**2.47**	**2.38**	**2.30**	**2.21**	**2.15**	**2.08**	**2.04**	**1.98**	**1.94**	**1.91**	
36	4.11	3.26	2.86	2.63	2.48	2.36	2.28	2.21	2.15	2.10	2.06	2.03	1.98	1.93	1.87	1.82	1.78	1.72	1.69	1.65	1.62	1.59	1.56	1.55	36
	7.39	**5.25**	**4.38**	**3.89**	**3.58**	**3.35**	**3.18**	**3.04**	**2.94**	**2.86**	**2.78**	**2.72**	**2.62**	**2.54**	**2.43**	**2.35**	**2.26**	**2.17**	**2.12**	**2.04**	**2.00**	**1.94**	**1.90**	**1.87**	
38	4.10	3.25	2.85	2.62	2.46	2.35	2.26	2.19	2.14	2.09	2.05	2.02	1.96	1.92	1.85	1.80	1.76	1.71	1.67	1.63	1.60	1.57	1.54	1.53	38
	7.35	**5.21**	**4.34**	**3.86**	**3.54**	**3.32**	**3.15**	**3.02**	**2.91**	**2.82**	**2.75**	**2.69**	**2.59**	**2.51**	**2.40**	**2.32**	**2.22**	**2.14**	**2.08**	**2.00**	**1.97**	**1.90**	**1.86**	**1.84**	
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.07	2.04	2.00	1.95	1.90	1.84	1.79	1.74	1.69	1.66	1.61	1.59	1.55	1.53	1.51	40
	7.31	**5.18**	**4.31**	**3.83**	**3.51**	**3.29**	**3.12**	**2.99**	**2.88**	**2.80**	**2.73**	**2.66**	**2.56**	**2.49**	**2.37**	**2.29**	**2.20**	**2.11**	**2.05**	**1.97**	**1.94**	**1.88**	**1.84**	**1.81**	
42	4.07	3.22	2.83	2.59	2.44	2.32	2.24	2.17	2.11	2.06	2.02	1.99	1.94	1.89	1.82	1.78	1.73	1.68	1.64	1.60	1.57	1.54	1.51	1.49	42
	7.27	**5.15**	**4.29**	**3.80**	**3.49**	**3.26**	**3.10**	**2.96**	**2.86**	**2.77**	**2.70**	**2.64**	**2.54**	**2.46**	**2.35**	**2.26**	**2.17**	**2.08**	**2.02**	**1.94**	**1.91**	**1.85**	**1.80**	**1.78**	
44	4.06	3.21	2.82	2.58	2.43	2.31	2.23	2.16	2.10	2.05	2.01	1.98	1.92	1.88	1.81	1.76	1.72	1.66	1.63	1.58	1.56	1.52	1.50	1.48	44
	7.24	**5.12**	**4.26**	**3.78**	**3.46**	**3.24**	**3.07**	**2.94**	**2.84**	**2.75**	**2.68**	**2.62**	**2.52**	**2.44**	**2.32**	**2.24**	**2.15**	**2.06**	**2.00**	**1.92**	**1.88**	**1.82**	**1.78**	**1.75**	
46	4.05	3.20	2.81	2.57	2.42	2.30	2.22	2.14	2.09	2.04	2.00	1.97	1.91	1.87	1.80	1.75	1.71	1.65	1.62	1.57	1.54	1.51	1.48	1.46	46
	7.21	**5.10**	**4.24**	**3.76**	**3.44**	**3.22**	**3.05**	**2.92**	**2.82**	**2.73**	**2.66**	**2.60**	**2.50**	**2.42**	**2.30**	**2.22**	**2.13**	**2.04**	**1.98**	**1.90**	**1.86**	**1.80**	**1.76**	**1.72**	
48	4.04	3.19	2.80	2.56	2.41	2.30	2.21	2.14	2.08	2.03	1.99	1.96	1.90	1.86	1.79	1.74	1.70	1.64	1.61	1.56	1.53	1.50	1.47	1.45	48
	7.19	**5.08**	**4.22**	**3.74**	**3.42**	**3.20**	**3.04**	**2.90**	**2.80**	**2.71**	**2.64**	**2.58**	**2.48**	**2.40**	**2.28**	**2.20**	**2.11**	**2.02**	**1.96**	**1.88**	**1.84**	**1.78**	**1.73**	**1.70**	

Table II. 5% (Roman Type) and 1% (Bold Face Type) Points for the Distribution of F

n_2	n_1 degrees of freedom (for greater mean square)																								n_2
	1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	24	30	40	50	75	100	200	500	∞	
50	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	2.07	2.02	1.98	1.95	1.90	1.85	1.78	1.74	1.69	1.63	1.60	1.55	1.52	1.48	1.46	1.44	50
	7.17	**5.06**	**4.20**	**3.72**	**3.41**	**3.18**	**3.02**	**2.88**	**2.78**	**2.70**	**2.62**	**2.56**	**2.46**	**2.39**	**2.26**	**2.18**	**2.10**	**2.00**	**1.94**	**1.86**	**1.82**	**1.76**	**1.71**	**1.68**	
55	4.02	3.17	2.78	2.54	2.38	2.27	2.18	2.11	2.05	2.00	1.97	1.93	1.88	1.83	1.76	1.72	1.67	1.61	1.58	1.52	1.50	1.46	1.43	1.41	55
	7.12	**5.01**	**4.16**	**3.68**	**3.37**	**3.15**	**2.98**	**2.85**	**2.75**	**2.66**	**2.59**	**2.53**	**2.43**	**2.35**	**2.23**	**2.15**	**2.06**	**1.96**	**1.90**	**1.82**	**1.78**	**1.71**	**1.66**	**1.64**	
60	4.00	3.15	2.76	2.52	2.37	2.25	2.17	2.10	2.04	1.99	1.95	1.92	1.86	1.81	1.75	1.70	1.65	1.59	1.56	1.50	1.48	1.44	1.41	1.39	60
	7.08	**4.98**	**4.13**	**3.65**	**3.34**	**3.12**	**2.95**	**2.82**	**2.72**	**2.63**	**2.56**	**2.50**	**2.40**	**2.32**	**2.20**	**2.12**	**2.03**	**1.93**	**1.87**	**1.79**	**1.74**	**1.68**	**1.63**	**1.60**	
65	3.99	3.14	2.75	2.51	2.36	2.24	2.15	2.08	2.02	1.98	1.94	1.90	1.85	1.80	1.73	1.68	1.63	1.57	1.54	1.49	1.46	1.42	1.39	1.37	65
	7.04	**4.95**	**4.10**	**3.62**	**3.31**	**3.09**	**2.93**	**2.79**	**2.70**	**2.61**	**2.54**	**2.47**	**2.37**	**2.30**	**2.18**	**2.09**	**2.00**	**1.90**	**1.84**	**1.76**	**1.71**	**1.64**	**1.60**	**1.56**	
70	3.98	3.13	2.74	2.50	2.35	2.23	2.14	2.07	2.01	1.97	1.93	1.89	1.84	1.79	1.72	1.67	1.62	1.56	1.53	1.47	1.45	1.40	1.37	1.35	70
	7.01	**4.92**	**4.08**	**3.60**	**3.29**	**3.07**	**2.91**	**2.77**	**2.67**	**2.59**	**2.51**	**2.45**	**2.35**	**2.28**	**2.15**	**2.07**	**1.98**	**1.88**	**1.82**	**1.74**	**1.69**	**1.62**	**1.56**	**1.53**	
80	3.96	3.11	2.72	2.48	2.33	2.21	2.12	2.05	1.99	1.95	1.91	1.88	1.82	1.77	1.70	1.65	1.60	1.54	1.51	1.45	1.42	1.38	1.35	1.32	80
	6.96	**4.88**	**4.04**	**3.56**	**3.25**	**3.04**	**2.87**	**2.74**	**2.64**	**2.55**	**2.48**	**2.41**	**2.32**	**2.24**	**2.11**	**2.03**	**1.94**	**1.84**	**1.78**	**1.70**	**1.65**	**1.57**	**1.52**	**1.49**	
100	3.94	3.09	2.70	2.46	2.30	2.19	2.10	2.03	1.97	1.92	1.88	1.85	1.79	1.75	1.68	1.63	1.57	1.51	1.48	1.42	1.39	1.34	1.30	1.28	100
	6.90	**4.82**	**3.98**	**3.51**	**3.20**	**2.99**	**2.82**	**2.69**	**2.59**	**2.51**	**2.43**	**2.36**	**2.26**	**2.19**	**2.06**	**1.98**	**1.89**	**1.79**	**1.73**	**1.64**	**1.59**	**1.51**	**1.46**	**1.43**	
125	3.92	3.07	2.68	2.44	2.29	2.17	2.08	2.01	1.95	1.90	1.86	1.83	1.77	1.72	1.65	1.60	1.55	1.49	1.45	1.39	1.36	1.31	1.27	1.25	125
	6.84	**4.78**	**3.94**	**3.47**	**3.17**	**2.95**	**2.79**	**2.65**	**2.56**	**2.47**	**2.40**	**2.33**	**2.23**	**2.15**	**2.03**	**1.94**	**1.85**	**1.75**	**1.68**	**1.59**	**1.54**	**1.46**	**1.40**	**1.37**	
150	3.91	3.06	2.67	2.43	2.27	2.16	2.07	2.00	1.94	1.89	1.85	1.82	1.76	1.71	1.64	1.59	1.54	1.47	1.44	1.37	1.34	1.29	1.25	1.22	150
	6.81	**4.75**	**3.91**	**3.44**	**3.14**	**2.92**	**2.76**	**2.62**	**2.53**	**2.44**	**2.37**	**2.30**	**2.20**	**2.12**	**2.00**	**1.91**	**1.83**	**1.72**	**1.66**	**1.56**	**1.51**	**1.43**	**1.37**	**1.33**	
200	3.89	3.04	2.65	2.41	2.26	2.14	2.05	1.98	1.92	1.87	1.83	1.80	1.74	1.69	1.62	1.57	1.52	1.45	1.42	1.35	1.32	1.26	1.22	1.19	200
	6.76	**4.71**	**3.88**	**3.41**	**3.11**	**2.90**	**2.73**	**2.60**	**2.50**	**2.41**	**2.34**	**2.28**	**2.17**	**2.09**	**1.97**	**1.88**	**1.79**	**1.69**	**1.62**	**1.53**	**1.48**	**1.39**	**1.33**	**1.28**	
400	3.86	3.02	2.62	2.39	2.23	2.12	2.03	1.96	1.90	1.85	1.81	1.78	1.72	1.67	1.60	1.54	1.49	1.42	1.38	1.32	1.28	1.22	1.16	1.13	400
	6.70	**4.66**	**3.83**	**3.36**	**3.06**	**2.85**	**2.69**	**2.55**	**2.46**	**2.37**	**2.29**	**2.23**	**2.12**	**2.04**	**1.92**	**1.84**	**1.74**	**1.64**	**1.57**	**1.47**	**1.42**	**1.32**	**1.24**	**1.19**	
1000	3.85	3.00	2.61	2.38	2.22	2.10	2.02	1.95	1.89	1.84	1.80	1.76	1.70	1.65	1.58	1.53	1.47	1.41	1.36	1.30	1.26	1.19	1.13	1.08	1000
	6.66	**4.62**	**3.80**	**3.34**	**3.04**	**2.82**	**2.66**	**2.53**	**2.43**	**2.34**	**2.26**	**2.20**	**2.09**	**2.01**	**1.89**	**1.81**	**1.71**	**1.61**	**1.54**	**1.44**	**1.38**	**1.28**	**1.19**	**1.11**	
∞	3.84	2.99	2.60	2.37	2.21	2.09	2.01	1.94	1.88	1.83	1.79	1.75	1.69	1.64	1.57	1.52	1.46	1.40	1.35	1.28	1.24	1.17	1.11	1.00	∞
	6.64	**4.60**	**3.78**	**3.32**	**3.02**	**2.80**	**2.64**	**2.51**	**2.41**	**2.32**	**2.24**	**2.18**	**2.07**	**1.99**	**1.87**	**1.79**	**1.69**	**1.59**	**1.52**	**1.41**	**1.36**	**1.25**	**1.15**	**1.00**	

TABLE III. TABLE OF χ^2 PROBABILITY SCALE (FROM R. A. FISHER'S TABLE)

Degrees of Freedom	*Chance of Exceeding Given Value of χ^2*							*Degrees of Freedom*
	.50	.30	.20	.10	.05	.02	.01	
n	*Values of χ^2*							*n*
1	.45	1.07	1.64	2.71	3.84	5.41	6.63	1
2	1.39	2.41	3.22	4.60	5.99	7.82	9.21	2
3	2.37	3.66	4.64	6.25	7.81	9.84	11.34	3
4	3.36	4.88	5.99	7.78	9.49	11.67	13.28	4
5	4.35	6.06	7.29	9.24	11.07	13.39	15.09	5
6	5.35	7.23	8.56	10.64	12.59	15.03	16.81	6
7	6.35	8.38	9.80	12.02	14.07	16.62	18.47	7
8	7.34	9.52	11.03	13.36	15.51	18.17	20.09	8
9	8.34	10.66	12.24	14.68	16.92	19.68	21.67	9
10	9.34	11.78	13.44	15.99	18.31	21.16	23.21	10
11	10.34	12.90	14.63	17.27	19.67	22.62	24.72	11
12	11.34	14.01	15.81	18.55	21.03	24.05	26.22	12
13	12.34	15.12	16.98	19.81	22.36	25.47	24.69	13
14	13.34	16.22	18.15	21.06	23.68	26.87	29.14	14
15	14.34	17.32	19.31	22.31	25.00	28.26	30.58	15
16	15.34	18.42	20.46	23.54	26.30	29.63	32.00	16
17	16.34	19.51	21.61	24.77	27.59	30.99	33.41	17
18	17.34	20.60	22.76	25.99	28.87	32.35	34.80	18
19	18.34	21.69	23.90	27.20	30.14	33.69	36.19	19
20	19.34	22.77	25.04	28.41	31.41	35.02	37.57	20
21	20.34	23.86	26.17	29.61	32.67	36.34	38.93	21
22	21.34	24.94	27.30	30.81	33.92	37.66	40.29	22
23	22.34	26.02	28.43	32.01	35.17	38.97	41.64	23
24	23.34	27.10	29.55	33.20	36.41	40.27	42.98	24
25	24.34	28.17	30.67	34.38	37.65	41.57	44.31	25
26	25.34	29.25	31.79	35.56	38.88	42.86	45.64	26
27	26.34	30.32	32.91	36.74	40.11	44.14	46.96	27
28	27.34	31.39	34.03	37.82	41.34	45.42	48.28	28
29	28.34	32.46	35.14	39.09	42.56	46.69	49.59	29
30	29.34	33.53	36.25	40.26	43.77	47.96	50.89	30

For larger values of n, $\sqrt{2\chi^2} - \sqrt{2n-1}$ may be referred approximately to normal probability scale.

INDEX